THE UNIVERSITY
of EDINBURGH

Exploring the interactome of the *Drosophila* SUMO E3 ligase PIAS

Emily Fowler

A thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

2019

Declaration

I declare that I alone composed this thesis and that the work presented here is my own unless otherwise stated. This work has not been submitted for any other degree or personal qualification.

Emily R. Fowler
June 2019

Abstract

Drosophila Protein Inhibitor of Activated STAT (PIAS) is required for chromatin organisation. PIAS proteins are SUMO E3 ligases that provide target specificity for the addition of small ubiquitin like modifiers (SUMO) to proteins to alter their function.

Previous experiments in the Heun laboratory, which used a PIAS overexpression system, identified a relationship between PIAS and XNP (X-linked nuclear protein). Both proteins play a role in silencing transposable elements and likely interact. XNP and ADD1 (ADD (ATRX-DNMT3-DNMT3L) domain-containing protein 1) together make up the two halves of their mammalian homologue, the chromatin remodeller ATRX. XNP and ADD1 have been shown by us and others to be SUMOylated. We hypothesise that the SUMOylation of XNP by PIAS could be essential for its role in chromatin organisation and aimed to study this interaction within a *Drosophila* model using S2 tissue culture cells and embryos.

In this thesis, I defined the interactomes of PIAS, ADD1 and XNP using immunoprecipitation-mass spectrometry (IP-MS) in the presence and absence of SUMOylation. Using CRISPR/Cas9 epitope tagged PIAS in S2 cell lines and primary antibodies against endogenous PIAS, I could confirm an interaction between PIAS and XNP in a SUMO dependent manner. Interestingly, I found PIAS interacts with ADD1 and heterochromatin protein 1 (HP1) in a SUMO independent manner. These interactomes support the hypothesis that SUMOylation of XNP by PIAS is involved in the association of XNP with ADD1 to promote heterochromatin organisation.

Lay Summary

The genome is made of DNA and contains all the instructions necessary to create an entire organism. These instructions are encoded in the form of genes and when read, control how a cell behaves. A copy of the genome is stored in every cell. However, if all the DNA in a single cell were to be stretched out, it would be two meters long; therefore, DNA is packaged. This packaging also serves a role in regulating the reading of the genes. Loose packaging is proposed to promote reading, while tight packaging leads to silencing of genes. In some cancers and diseases the packaging can go wrong, meaning that genes that should be silent are left on, allowing cells to grow uncontrollably. Through understanding the way DNA packaging is regulated further cancer treatments may be developed.

Many proteins regulate DNA packaging. My research has looked at PIAS, which is a protein that has the ability to modify other proteins by attaching an additional small protein called SUMO. In this thesis I describe my work to see which proteins interact with PIAS and how their activity might be controlled by the addition of SUMO.

.

Acknowledgements

I would like to thank the BBSRC for funding my position as part of the EASTBIO programme and Sara Buonomo for helping me to secure this, which brought me to Edinburgh for my PhD. Edinburgh has been fantastic place for seminars, courses and training, with an excellent postgraduate writing programme that helped me to enjoy writing-up. I have met many remarkable scientists during my time in Edinburgh and I am grateful to all the people that I have discussed my work with over the last four years who have provided guidance, encouragement and importantly have helped to improve my work, understanding and communication.

I thank Patrick and Atlanta for supervision. Thanks to Patrick for giving me this interesting project to work on and to everyone who has worked on *Drosophila* PIAS over the last eighteen years allowing me to piece together how this protein may be acting at chromatin and put what is known about PIAS into the context of our current understandings of heterochromatin organisation. I thank Atlanta for her help over the last two and a half years and for helping me to think more clearly about the experiments I performed and how I communicated the results, especially in writing this thesis. I also thank my thesis committee members, Kevin Hardwick and David Leach for comments and discussions on the direction of my work - through changes in questions and direction as my understanding of the research field, my capabilities and possibilities became apparent, ensuring that I had the content and time to write a thesis.

Thanks to Matt and Liz in the EPPF, everyone in the PhD students office, past and present members of the Cook and Heun labs and to Dr. Christos Spanos for running all my mass spec samples. I would also like to thank Matilde Fabe for taking over the project and her enthusiasm in discussing the future direction of this work.

Personally, I would not have been able to complete my PhD without the grounding provided to me during my undergraduate degree at the University of York. I am grateful to Dawn and Betsy for introducing me to the research environment and for mentoring me at that time. Lastly and most importantly I thank my family: my partner, Chris, for everything over the last four years and for supporting me through this PhD and my parents for providing me with opportunities to keep learning and thinking about what is important to me.

Contents

Appendices

List of Figures

Appendix Figures

List of Tables

Appendix Tables

Abbreviations

μg	$1x10^{-6}$ grams
μL	$1x10^{-6}$ litres
$A_{280\ (254)\ (260)\ (595)}$	absorbance at 280 (254, 260, 280) nanometers
aa	amino acid (s)
ADD1	ATRX-DMNT1-DMNT3
Amp	ampicillin
APS	ammonium persulfate
ATP	adenosine triphosphate
ATRX	α thalassemia/mental retardation syndrome X-linked protein
bacmid	bacterial artificial chromosome
BEAF-32	boundary element associated factor-32
bp	base pair
BSA	bovine serum albumin
Cas9	CRISPR-associated protein 9
cDNA	complementary deoxyribonucleic acid
ChIP	chromatin immunoprecipitation
CNS	central nervous system
CRISPR	clustered regularly interspaced short palindromic repeats
DAXX	death domain associated protein
ddH_2O	double distilled water
DLS	dynamic light scattering
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DNMT	DNA methyltransferase
dNTP	deoxyribonucleoside triphosphate
ECL	enhanced chemiluminescence
EDTA	ethylene diamine-tetra acetic acid
e(var)	enhancer of variegation gene
FDR	false discovery rate
(nano)g	(nano)grams
x g	relative centrifugal force
GFP	green-fluorescent protein
GST	glutathione S-transferase
hr	hour(s)
H3/H3.3	histone H3/ histone H3.3
HAT	histone acetyltransferase

HDAC	histone deacetylase
HIRA	histone regulator A
HDRT	homology directed repair template
HP1	heterochromatin protein 1
IF	immunofluorescence
Ig	immunoglobulin
IP	immunoprecipitation
IPTG	isopropyl β-D-1-thiogalactopyranoside
kb	kilo-base pair
kDa	kilo-Dalton
KMT	lysine methyltransferase
KDM	lysine demethylase
KO	knockout
m(L)	(milli)litres
LFQ	label free quantification
M	molar
Mb	mega-base pair
MQ	milli-Q water
MS	mass spectrometry
min	minute(s)
NEM	N-ethylmaleimide
nt	nucleotide
OD_{600}	optical density at 600 nanometers
ORF	open reading frame
PABP	polyadenylate binding protein
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEV	position effect variegation
pH	-log10[H+]
PIAS	Protein Inhibitor of Activated STAT protein
Piwi	P-element Induced WImpy testes
PTGS	post-transcriptional gene silencing
PTM	post-translational modification
RISC	RNA induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
RNAPol II	RNA polymerase II
RNAse	ribonuclease

row (*row*)	relative of woc
rpm	revolutions per minute
RT	room temperature
s	second(s)
SDS	sodium dodecyl sulphate
SMT3	suppressor of MIF Two Nr. 3
STAT	signal transducer and activator of transcription
SUMO	small ubiquitin-like modifier
su(var)	suppressor of variegation gene
TADs	topologically associated domains
Tafs	transcription associated factors
UBLs	ubiquitin-like modifiers
UV	ultraviolet
V	volts
woc (*woc*)	without children
wt	wild type
XNP	X-linked nuclear protein
YFP	yellow-fluorescent protein
℃	degree celcius

SUMOylation, similarly to ubiquitination, is involved in all stages of a cell's biology and regulation. Although this thesis in large part aims to investigate the interactome of the SUMO E3 ligase PIAS, it will focus on the role of PIAS SUMOylation in heterochromatin formation and maintenance. Therefore this thesis takes the aim of identifying potential mechanisms of PIAS's role through these four pathways which have been identified to interact with PIAS:

1. the chromatin remodeller ATRX (ADD1 and XNP in *Drosophila*)
2. heterochromatin- euchromatin boundary regulators BEAF-32 and other insulator proteins
3. heterochromatin establishment through the piRNA pathway
4. heterochromatin establishment through the histone methyltransferase Eggless.

Due to the many and varied roles of SUMOylation this thesis will begin by reviewing large areas of biology and the role of SUMOylation. It will then continue onto results, which will specifically investigate the interactome of PIAS, before attempting to tie together biological theory and data within the conclusions and perspectives, finishing with hypothetical models.

Chapter 1 General Introduction

1.1 Genome Definition and Content

The genome comprises the total genetic information of an organism, in the form of DNA, and is packaged into the cell nucleus. The size of the genome varies depending on the organism, with a general trend that eukaryotic genomes are larger than prokaryotic genomes and vertebrate genomes are larger than invertebrate genomes. The human genome consists of 3.1 giga-bases (gb) of DNA packaged into 23 chromosomes. For comparison, the *Drosophila melanogaster* genome contains 180 mega-bases (mb) in 4 chromosomes.

1.2 Central Dogma of Molecular Biology

The genome is split into non-coding and coding regions. Coding regions of DNA are known as genes which are transcribed and translated into functional units in the form of RNA and proteins. Proteins are the main structural, functional and regulatory units in the cell. The proteins in a cell at any one time make up the proteome of the cell and this is dynamically regulated through control at the DNA, RNA and protein level.

1.3 Transcription and Processing of mRNAs

In eukaryotes, RNA Polymerase II (RNAPol II) is responsible for transcribing messenger RNA (mRNA) that codes for proteins, for this the TATA binding protein (TBP) must recognise the DNA upstream of gene start sites, signified by the DNA sequence TATA. This recognition within the highly packaged genome is aided by transcription associated factors (Tafs)[1].

After gene transcription, precursor mRNA (pre-mRNA) contains exons (coding directly for protein) and introns (not coding for protein). Introns are removed and exons are spliced together to generate a mature mRNA, which is capped and poly-adenylated. Post processing, mRNA is exported from the nucleus and translated into protein.

1.4 Genome Organisation

DNA is packaged inside the cell nucleus in the form of chromatin. The basic repeating unit of chromatin is the nucleosome, which is made up of 147bp of DNA wrapped around eight histone proteins (two histone 3 (H3) and two histone 4 (H4) dimers flanked by two heterodimers of histone 2A (H2A) and histone 2B (H2B))[2,3,4]. This structure is referred to as “beads-on-a-string”. Nucleosomes are further packaged in a series of folding steps regulated by protein-DNA interactions (Figure 1–1).

Chromatin attaches to an underlying structure in the nucleus. Biochemical extraction methods first defined a filamentous ribo-proteinaceous nuclear matrix[5] which provides a site for DNA replication[6], transcription[7], splicing[8] and chromatin remodelling[9]. More recently an idea of a more dynamic “nuclear mesh” has emerged (reviewed in Nozawa 2019[10]). Chromatin-associated RNAs form ‘RNA clouds’ over specific clusters of active gene promoters and their distal enhancers which loop into proximity (reviewed in Li, 2019)[11].

3D genome organisation (genome structure) has emerged as an area of great interest with relation to gene transcription (genome function). Spatially certain regions of the genome are in greater contact with each other than with other regions and form distinct subcompartments, known as topologically associated domains (TADs)[12]. However, the functional significance of these domains across evolutionally divergent organisms is under review with increased resolution providing new insights into chromatin interactions within TADs, favouring a model[13] based on chromatin states[14]. These chromatin states may change with transcriptional changes[15], RNA Pol II occupancy and an increase in DNA accessibility at boundary regions. These boundary regions are enriched for architectural proteins, which bind specific DNA sequences such as CCCTC-binding factor (CTCF) which has insulator activity[16,17].

Principles of protein clustering into “factories” containing high local concentrations of factors required for specific cellular process is not a new concept[18,19]. However, new interest in chromatin self-organisation based on its biophysical properties has given rise to concepts of phase-separated subcompartments. Soluble proteins can bind to chromatin scaffolds and each other, oligomerising, bridging or cross-linking to organise chromatin domains with particular properties. Homotypic attraction between similar chromatin states can aid enhancer-promoter contacts (reviewed in Robson *et al.,* 2019[20]). Alternatively, multivalent interactions can drive the assembly of liquid-like dynamic nuclear bodies around their chromatin binding sites. These can coalesce and are in constant and rapid exchange with the surrounding nucleoplasm. In addition, *in vitro* intrinsically disordered regions of proteins interact with RNA and form liquid droplets. The size, density, and stability of chromatin bodies depends on the concentration and interactions of the molecules involved[21]. Sequence-specific interactions may define nucleation sites, but feedback-based mechanisms may be required to reinforce a given chromatin state.

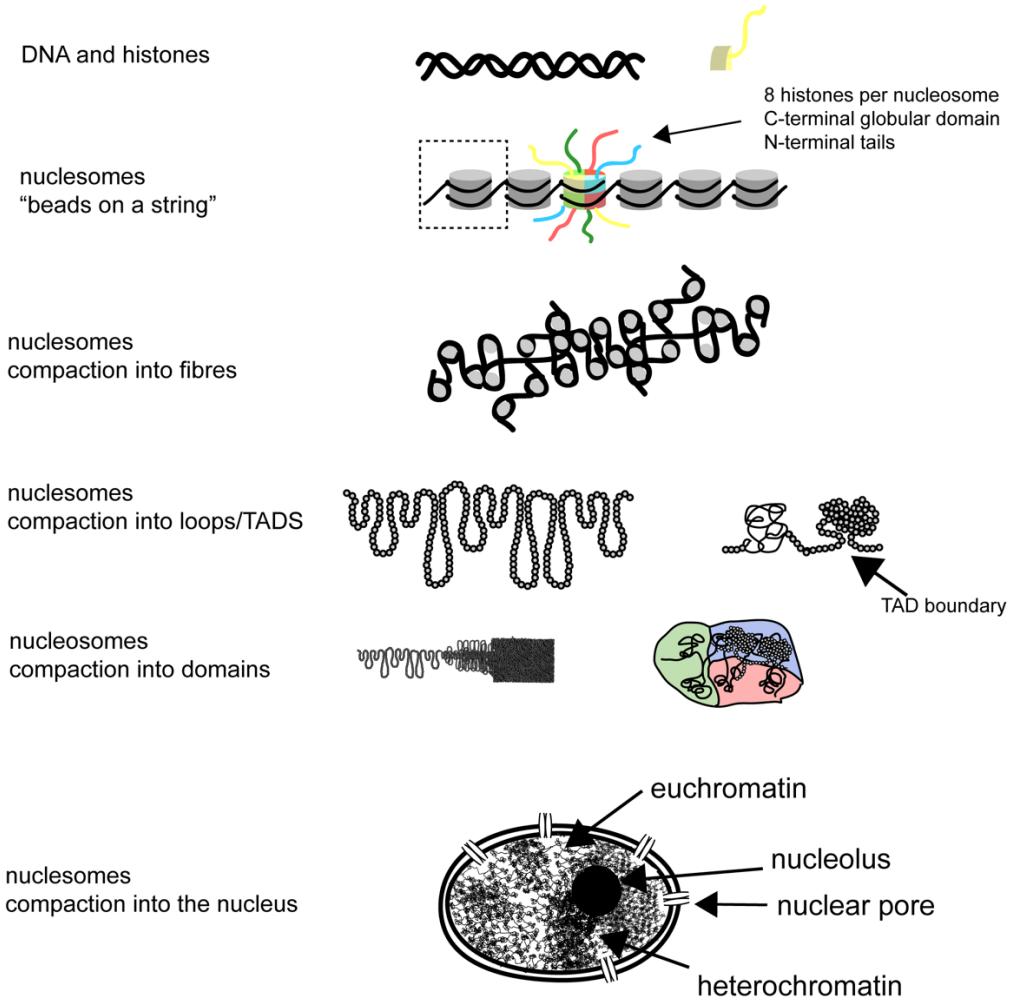


Figure 1–1 DNA organisation within the cell nucleus

DNA is packaged into nucleosomes that are organised into “beads on a string” chromatin structures. Chromatin is dynamically packaged inside the cell nucleus, with increasing levels of compaction. Figure adapted from Hansen, 2018[22].

1.5 Chromatin Modifications

Chromatin can be defined by its state. Histone N-terminal tails can be modified by the addition of post-translational modifications (PTMs) to provide a "histone code". Specific PTMs or combinations of PTMs are "read" by proteins that have specific chromatin binding domains[23] [24] and as a consequence, chromatin may be modified or remodelled to promote or restrict access to the underlying genes.

Two well characterised histone PTMs are acetylation (which is generally associated with active chromatin) and methylation (which is generally associated with repressive chromatin) of lysine residues, which can be mono-, di- or tri-methylated[25]. Addition and removal of histone PTMs is regulated through enzymes and protein complexes that act as "writers" and "erasers" of chromatin marks. For acetylation, histone acetyltransferases (HATs) deposits acetyl groups, and histone deacetylases (HDACs) catalyse their removal. For methylation, lysine methyltransferases (KMTs) and lysine demethylases (KDMs) add and remove methyl groups respectively. PTMs can alter the interaction between histones and DNA[26], which influences chromatin structure and gene expression either by blocking or activating protein recruitment or enhancing or inhibiting interactions. The addition of modifications that are "above the DNA" and can be stably inherited over generations are called epigenetic.

1.6 DNA Modifications

DNA can be modified by *de novo* methylation of its bases; this is carried out by two DNA methyltransferases (DNMTs): DNMT3a and DNMT3b[27]. Both are active during gametogenesis and re-implantation[28], acting to silence regions of the genome[29]. The most prominent DNA modification is the addition of a methyl group at the C5 position of the pyrimidine ring of the cytosine residue (5meC) in the context of CG dinucleotides. In vertebrate genomes ~70% of CG dinucleotides are methylated[30] and unmethylated CG dinucleotides (CpG islands) are typically found within gene promoters[31]. 5meC may exert a repressive effect on transcription by blocking association of transcription factors with promoters or blocking access to chromatin remodelling complexes and/or histone modifying enzymes. Proteins with a methyl binding domain (MBD) such as methyl CpG binding protein 2 (MeCP2)[32] bind

methylated DNA and recruit other proteins that act as chromatin regulators, including SIN3, HDAC1[33], ATRX[34] which act in a transcriptional repressor complex[35]. Gene silencing is established through a cross-talk between DNA methylation and histone modification. Another example of this is with the DMNT proteins. In addition to their role as *de novo* methyl transferases, the ATRX-DNMT3-DNMT3L (ADD) domain of DMNT3 binds histone deacetylation complex 1 (HDAC1)[36], MeCP2[37] and polycomb repressive complex 2 (Pc2)[38,39], forming part of a larger repressive complex.

Drosophila melanogaster have low levels of DNA methylation [40,41], with only one *de novo* DMNT homolog[42] which is non-essential[43]. Histone modifications however, play a key role in early embryonic epigenetic reprogramming[44], perhaps indicative of different underlying mechanisms of regulation of gene expression between vertebrates and invertebrates.

1.7 Histone Variants

Although most nucleosomes consist of the canonical histones: H2A, H2B, H3 and H4; histone proteins can differ in their amino acid sequence. H2A and H3 histone variants alter the biophysical properties of the nucleosome[45] altering the structure and function of chromatin, which influences chromatin interaction partners. H3 histone variants have been well-characterised, one of which is H3.3. Histone H3.3 differs from canonical H3 by five amino acid residues. H3.3 is found at telomeric[46] and pericentric heterochromatin repeats[47], intergenic regions and introns[48] and plays an essential role in chromatin stability[49].

1.8 Histone 3.3

In *Drosophila*, H3 variants can compensate for each other in euchromatic regions. At highly transcribed genes, in euchromatin, H3.3 is deposited into nucleosomes in a replication-independent manner by the histone regulator A (HIRA) complex to maintain genome accessibility. H3.3 could account for the epigenetic memory of an activated state[50]. In *Caenorhabditis elegans*, H3.3 is non-essential yet under heat shock, in H3.3 knockout animals, expression of heat shock genes is reduced and fewer animals survive[49].

H3.3 is also enriched at telomeres and the endogenous retrovirus IAP element[51], with incorporation at these sites mediated by DAXX/ATRX (see sections 1.12 and 2.4.3)[47]. In mouse embryonic stem cells H3.3 is essential, with its absence causing developmental retardation and embryonic lethality[52] and dense chromatin packaging[53].

1.9 Types of Chromatin

Chromatin was first described by its appearance under light microscopy[54] and at first was separated into two subtypes. Loosely packaged 'open' euchromatin was observed in the central region of the cell nucleus, whereas more densely packaged 'closed' heterochromatin was observed at the periphery, underlying the nuclear membrane and attached to the nuclear matrix surrounding the nucleolus[55]. Our current understanding of chromatin is more complex. Chromatin type is defined by combinations of histone variants, chromatin modifications, and association with chromatin binding proteins that lead to differential packaging/gene expression. In *Drosophila*, chromatin has been categorised into four[56], five[57], nine or even 30 different types[58,59].

1.10 Constitutive Heterochromatin Definition, Function and Importance

This thesis will focus on constitutive heterochromatin. Constitutive heterochromatin is enriched at centromeric, pericentromeric and subtelomeric regions of the genome that contain repetitive DNA sequences[60]. The histone methylation mark H3K9me3[61] and heterochromatin protein 1 (HP1)[62] define constitutive heterochromatin at the molecular level. Constitutive heterochromatin has essential roles in nuclear architecture, chromosome segregation, genome stability via stabilising recombination, and repression of gene expression and silencing of transposable elements (reviewed in Janssen et al, 2018[63]). Histone variants can also be modified by histone methyltransferases, with the introduction of H3.3K9me3 being important for heterochromatin formation at mouse telomeres[64]. In *Drosophila,* $H3.3^{K9R}$ mutants are sterile[65], demonstrating that H3.3K9 is essential for propagation of offspring.

1.11 Constitutive Heterochromatin Structure, Establishment and Maintenance

Heterochromatin establishment involves multiple cellular machineries including complexes involved in: transcription of DNA; RNA interference (RNAi); protein-protein interactions; protein-RNA interactions and PTMs which can act singularly or in a combinatorial and complementary manner[66,67](reviewed in Allshire and Madhani 2017[68]). In *Schizosaccharomyces pombe,* heterochromatin is formed through the action of the RNA-induced transcriptional silencing complex (RITS), which is targeted to endogenous RNA transcripts by small interfering RNAs (siRNAs) generated from the cleavage products of transcribed repeats[69,70,71]. The RITS associates with the SUV39 histone H3K9 methyltransferase to deposit H3K9me3 and recruit HP1 to establish heterochromatin domains. The mechanism of RITS silencing and heterochromatin domain formation is shown in Figure 1–2.

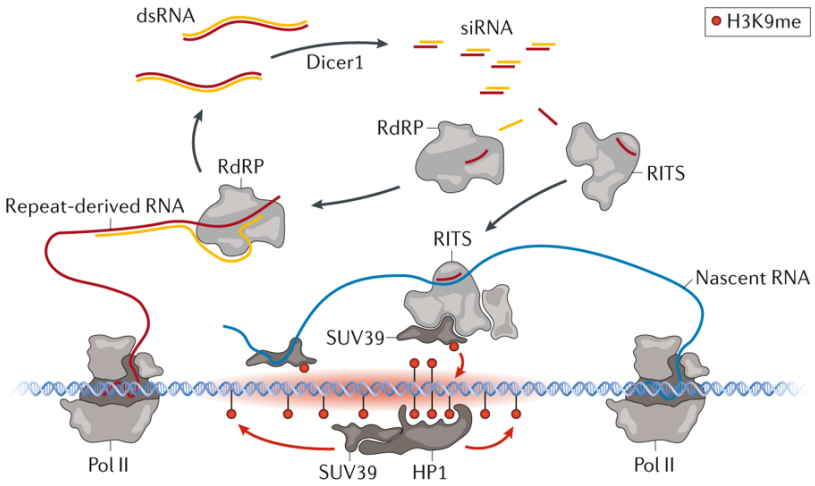


Figure 1–2 RNA-dependent and RNA-independent feedback loops for establishing and locally spreading heterochromatin in *Schizosaccharomyces pombe*

Fission yeast view of heterochromatin spreading. Figure taken from Li and Fu 2019[11] RNA is transcribed from RNA polymerase II (Pol II) and processed by RNA dependent RNA polymerase (RdRP) to form double stranded RNA (dsRNA). This dsRNA is then cleaved by Dicer 1 to form small interfering RNAs (siRNA). These siRNAs when bound to the RNA-induced transcriptional silencing complex (RITS) are targeted to complementary regions of the genome. Once targeted, the siRNA-RITS complex interacts with the methyltransferase enzyme (SUV39) to deposit H3K9me3 and recruit heterochromatin protein 1 (HP1), which through a positive feedback loop establishes a heterochromatin domain (red shadow).

Another RNAi pathway, the Piwi/piRNA pathway, acts to silence regions of the genome[72,73]. This pathway is conserved in metazoans[74] and is responsible for silencing transposable elements to prevent their mobilisation in germ cells[75,76,77] . This is an essential pathway in early embryos[78] and stem cells[79] (reviewed in Ozata *et al.,* 2019[80]). Piwi and piRNAs are imported into the nucleus to direct DNA methylation[81] of complementary repetitive elements to silence the endogenous retro-elements LINE 1 and IAP in mouse fetal testes[82] in pachytene sperm cells. Piwi and piRNAs are involved in epigenetic transcriptional silencing at imprinted genes[83] and initiate heterochromatin formation[72,84]. However, the molecular mechanisms of Piwi mediated repression, involving downstream proteins remains to be resolved. Piwi itself may interact with chromatin modifier proteins which deposit methyl marks at chromatin to repress transposable elements. The mechanism of Piwi mediated repression in *Drosophila melanogaster* is shown in Figure 1–3. In *Drosophila*, most primary piRNAs are generated in germline cells from pericentric and telomeric heterochromatic genomic loci (piRNA clusters) containing damaged repeated transposable element sequences. These primary piRNAs are amplified in a “Ping-Pong” cycle to generate secondary piRNAs. HP1 recruitment depends on the presence of small RNAs[85] and HP1 can also interact directly with Piwi[86,87].

It has been proposed that ATP-dependent RNA helicases, such as Rm62 and bel, may be involved in the production of *de novo* piRNAs and endogenous small interfering RNAs (endo-siRNAs) from euchromatic transposon insertion sites, provoking the activation of related small RNA pathways[88,89]. One of the other RNAi pathways in *Drosophila* involves a complex containing: AGO2 (an Argonaute family protein); Tudor-SN (a putative endonuclease (TSN)); VIG (the product of vasa intronic gene); fragile X mental retardation protein (Fmr1)[90,91]; and Rm62. Su(var)3-9 (histone methyltransferase)(SUV39 in Figure 1-3) was also found to act in this pathway[92]. Fmr1 and Tudor-SN also function as part of the Piwi/piRNA pathway, interacting with Piwi[93], demonstrating that proteins can be involved in multiple different silencing pathways.

The molecular details of the mechanisms of heterochromatin establishment and/or maintenance at pericentromeric and centromeric repeats in *Drosophila* remains unknown. By understanding mechanisms from multiple organisms it may be possible to unify concepts. Heterochromatin establishment is possibly aided by the siRNA pathway generating siRNA from major satellite repeats in complex with Argonaute proteins. In human cell lines, AGO2 interacts with the histone methyltransferase SETDB1 to aid heterochromatin spreading[94]. AGO2 mutant *Drosophila* embryos have defects in chromosome condensation, nuclear kinesis and spindle assembly, possibly correlated with defects in the formation of centromeric heterochromatin[95]. Therefore, regulation of AGO2 as part of the siRNA pathway plays an important role in heterochromatin assembly at these sites.

RNAs at the pericentromeres and telomeres act locally in *cis* and associate with histone lysine methyltransferases (KMTs)[67,96-98 99] to tri-methylate (me3) H3K9 introducing the repressive H3K9me3 mark[100]. HP1 is recruited to H3K9me3 through its “reader” N-terminal chromo domain[101,102]. HP1 self-associates through its C-terminal chromo shadow domain and recruits more KMT proteins in a positive feedback loop, causing heterochromatin spreading. HP1a proteins phase-separate in response to DNA binding or phosphorylation of their N-terminus to establish domain like structures[103]. In mammals HP1 also recruits HDACs and DNMTs, resulting in histone de-acetylation and DNA methylation[104-106]. Heterochromatin is maintained by different mechanisms to those used to establish it. HP1 acts as a platform to recruit factors to form silenced chromatin domains[107] and may serve as a maintenance factor for the heterochromatic state. The heterochromatin state is heritable through cell divisions and is maintained through epigenetic mechanisms involving “reader” and “writer” proteins[108].

In *Drosophila*, telomeres are maintained by successive transpositions of the retroelements: HeT-A and TART. This requires Spindle-E (an RNA helicase) and aub (an Argonaute protein)[109], both of which act in RNAi based mechanisms to maintain telomere elongation.

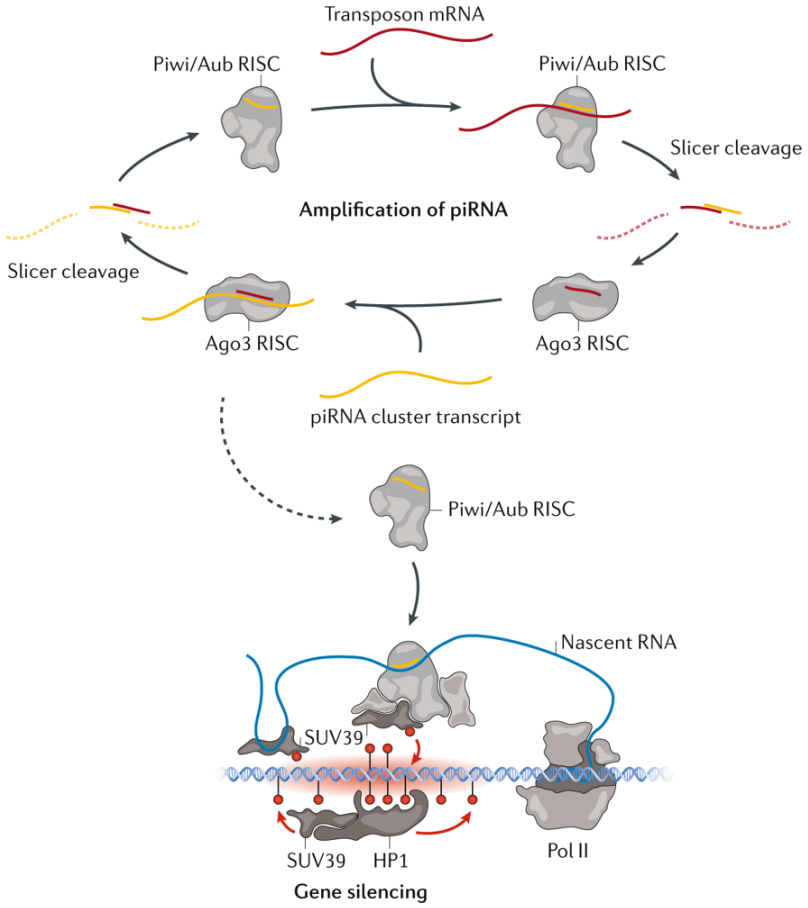


Figure 1–3 RNA-dependent and RNA-independent feedback loops for establishing and locally spreading heterochromatin in *Drosophila melanogaster*

Taken from Li and Fu 2019[11]. Transposon mRNA (top) is processed by the P-element Induced WImpy testis/Aubergine/ RNA-induced silencing complex (Piwi/Aub RISC) complex into Piwi-interacting RNAs (piRNAs). The “Ping-Pong” pathway of piRNA–directed RNA cleavage is proposed to amplify the piRNA population. It is catalyzed by the endonuclease (or “Slicer”) activities of the Piwi proteins Aubergine (Aub) and Argonaute3 (AGO3) to generate responder piRNAs. These responder piRNAs target nascent RNA transcribed from RNA polymerase II (PolII) to recruit the methyltransferase enzyme (SUV39) to deposit H3K9me3 and recruit heterochromatin protein 1 (HP1) to establish and spread, forming a heterochromatin domain (red shadow) resulting in gene silencing.

1.12 Chromatin Accessibility- Assembly and Remodelling

Chromatin remodellers can expose DNA through nucleosome sliding, nucleosome eviction or nucleosome unwrapping (by hydrolysis of ATP) to modulate histone-DNA contacts. This allows access of replication, transcription and repair machinery to the DNA[110]. Eukaryotes possess at least four ATP-dependent chromatin remodelling families: SWItching defective/Sucrose Non-Fermenting (SWI/SNF), Imitation SWItch (ISWI), Chromodomain-Helicase-DNA binding (NuRD/Mi-2/CHD), INOsitol requiring 80 (INO80) and SWRI[111]. The SWI/SNF family makes DNA accessible for transcription machinery and is important for regulation of gene expression[111,112]. The SNF2 family can be further divided into subfamilies as defined by phylogenetic comparison of the SNF2 ATPase region[113]. All SNF family chromatin remodellers contain a DNA/nucleosome binding domain and an ATP binding/ hydrolysis/ DNA translocation domain. One member of the SWI/SNF family is the ATRX protein.

Some chromatin remodellers have affinity for specific histone variants[114] and interact with histone chaperone proteins that are responsible for assembly of histone variants into nucleosomes. An example is the ATRX-DAXX-H3.3 complex. The histone chaperone death-domain-associated protein (DAXX) stabilises H3.3[115] and interacts with the chromatin remodeller ATRX to target DAXX to telomeres and pericentromeres where it deposits H3.3-containing nucleosomes[116]. DAXX can be post-translationally modified to aid interactions and complex formation with other repressive complexes. Disruption of the H3.3 pathway results in reduced levels of H3K9me3, ATRX and HP1α. This disruption in heterochromatin formation elevates chromosome instability at centromeres[117] and telomeres[64] resulting in detrimental effects on cell survival. See section 2.4.3 for further discussion of ATRX/DAXX/H3.3.

1.13 Chromatin States Define Cell Identity

The accessibility of the DNA to transcription factors is achieved through the interplay of chromatin modifications, chromatin remodellers and histone variants. This accessibility is influenced by the 3D organisation of the genome within the cell nucleus. This plays a key role in defining chromatin states that maintain cellular identity.

Chapter 2 Introduction

2.1 Heterochromatin

2.1.1 Heterochromatin Establishment and Maintenance in *Drosophila*

In *Drosophila,* early embryonic development is directed exclusively by maternal products (mRNA and proteins) deposited in the female gamete during oogenesis. *Drosophila* embryos undergo rapid cell division cycles, where only the nuclei divide and early cleavages take place without cytokinesis, these are therefore known as nuclear cycles (NC).

Drosophila lay eggs that can be collected and studied. A fertilised egg develops into an embryo. The progress of early *Drosophila* embryo development from NC1-14 is shown in Figure 2–1A. Separation of heterochromatin from euchromatin is initiated around NC10 and is completed by NC14[118]. This is visible by staining of the DNA (Figure 2-1B). The euchromatic state is defined by histone acetylation of H3K9 (H3K9ac) and is detectable in NC12[119] (Figure 2-1C). By NC13, nuclei adopt a Rabl configuration, heterochromatin is at the apical pole of the nuclei and is separated from euchromatin in the basal pole[85,120]. This configuration is thought to coincide with heterochromatinization[121], activation of widespread transcription from the zygotic genome[122], and cellularisation to enclose nuclei in the syncytial blastoderm marking the end of the nuclear cycle stages in development.

Heterochromatinization requires H3K4 demethylation by lysine-specific histone demethylase 1A (LSD1)[44]; H3K9 deacetylation by RPD3 (HDAC1)[44]; and methylation of H3K9 by the KMT Eggless[123]. Methylation of H3K9 is established in NC12, and increases which successive nuclear cycles to form stable chromatin domains by NC14[85,123]. HP1a is detectable in NC11[123] and by NC14 is stably associated with chromatin[124].

Seller *et al.,* 2019[123] show the histone methyltransferase Eggless, which is required in early embryos, is recruited to pericentric satellite DNA in NC10. However, the mechanism that recruits Eggless to chromatin remains unknown[123,125]. Seller *et al.,* 2019[123] hypothesise that maternally deposited small RNAs (including the 359bp satellite repeat, derived from sequences on the X-chromosome) could guide Eggless to transcripts. The initiation of heterochromatin formation could be triggered through an RNAi mediated mechanism as both the 359bp satellite repeat and Eggless are essential[85,123]. Eggless also plays a role in the piRNA pathway (discussed in section 1.11).

The Aravin lab has proposed that Su(var)2-10/PIAS, a SUMO E3 ligase, links the Piwi/piRNA pathway to the piRITS and that SUMOylation of the histone methyltransferase Eggless is required for its function in heterochromatin formation, as they demonstrate that gene repression by Eggless is dependent on Su(var)2-10[126]. It is possible that SUMOylation is required to “seed” heterochromatin formation as discussed in section 2.3.13 for HP1α targeting to pericentric heterochromatin. Seller *et al.,* 2019[123] suggest an alternative hypothesis that Eggless could bind a DNA binding protein to specify its localisation to specific satellite repeats[123]. It is possible that multiple mechanisms could act simultaneously.

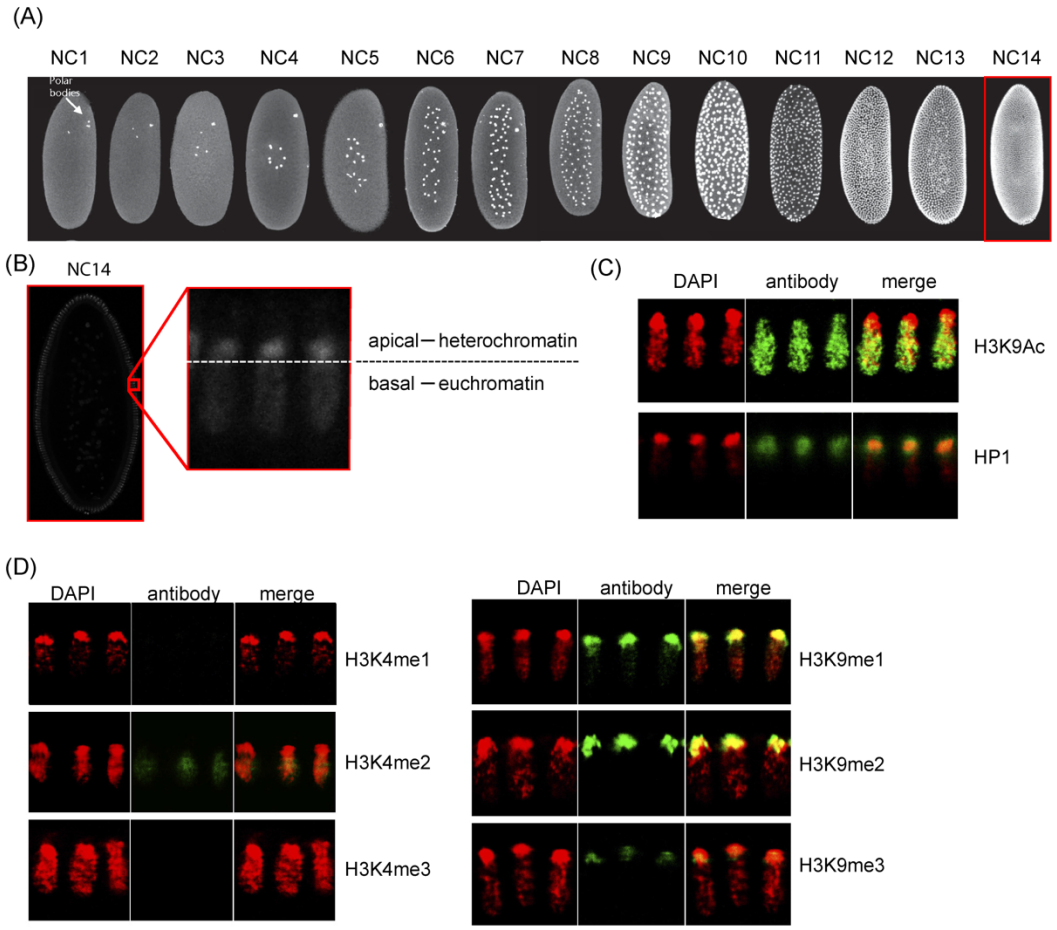


Figure 2–1 Visualisation of nucleic acid staining in *Drosophila melanogaster* embryos during development and formation of distinct euchromatin/heterochromatin domains within nuclear cycle 13/14 embryos distinguished by histone modifications and histone binding proteins.

(A) Images of DAPI stained embryos during the first 14 rounds of nuclear division (nuclear cycles 1-14) in a *Drosophila* embryo. Figure modified from Kotadia 2001[127]. **(B)** A *Drosophila* embryo in nuclear cycle 14 (NC 14), inset shows pre-blastoderm cells with the separation of heterochromatin and euchromatin in discrete domains by DAPI staining **(C)** Pre-blastoderm cells of a nuclear cycle 14 (NC14) wild-type *Drosophila* embryo, modified from Rudolph, 2007[44], immunostained with α-H2K9Ac and α-HP1 showing discrete staining in the basal and apical domain respectively. **(D)** Staining for histone modifications in pre-blastoderm cells of a NC14 *Drosophila* embryo, showing enrichment of H3K9 methylation at the apical domains.

2.1.2 Identification of Heterochromatin Proteins by Position-Effect Variegation Assays

Position-effect variegation (PEV) is an epigenetic phenomenon first described in *Drosophila* in 1930[128]. This was utilised to assay heterochromatin dynamics, for example in the In(1)wm^4 *Drosophila* line, whereby a chromosome inversion has placed the euchromatic white gene, that controls eye colour, adjacent to pericentric heterochromatin. Heterochromatin spreading into the white locus in some cells results in a shutdown of expression of the white gene, resulting in a white eye, whereas in other cells this silencing does not occur, leading to a red eye. This mosaic expression and clonal inheritance of the silenced state presents as a variegated phenotype in *Drosophila* eyes[129]. Combination of this with genetic knockouts can identify genes or DNA regions required for regulating heterochromatin dynamics through interpreting eye colour as an indicator of either spreading or recession of the heterochromatin boundary[130].

Around 150 proteins, most of which are structural or chromatin related, have been identified in this assay through introducing mutations in gene regions which give rise to altered chromatin composition[130]. Proteins encoded by these gene regions are classified as either suppressors or enhancers of variegation, Su(var)s or E(var)s respectively[131]. Su(var) proteins play a role in both heterochromatin establishment and maintenance.

The histone variant H3.3[132], histone deacetylases (e.g. Rpd35/HDAC1[133]), histone methyltransferases[134], phosphatases[135], components of the SUMOylation machinery, components of the RNAi machinery[70] (AGO2[95], Piwi[86,118], aubergine[118], Rm62[136] and vig[137], Fmr1[138]) and chromatin remodellers such as XNP[139] are all encoded by *Su(var)* genes. This thesis will focus on Su(var)205 - more commonly known as heterochromatin protein 1 (HP1)[140] and *Su(var)2-10* which encodes Su(var)2-10 also known as dPIAS or PIAS.

2.2 PIAS

2.2.1 The *Su(var)2-10* Gene Locus Encodes PIAS Proteins

The *Su(var)2-10* gene locus on chromosome 2R in *Drosophila* encodes Su(var)2-10 protein isoforms[131]. These isoforms share homology to vertebrate Protein Inhibitor of Activated STAT (PIAS) proteins[141], therefore Su(var)2-10 protein isoforms are also referred to in the literature as dPIAS proteins or just PIAS.

2.2.2 PIAS Expression and Conservation

PIAS is expressed ubiquitously in *Drosophila* with the highest levels of expression in the embryo[142]. In adults, expression is highest in the CNS and gonads[143]. At the cellular level, during *Drosophila* embryogenesis, PIAS localises to heterochromatin in nuclear cycles 10-14 coinciding with heterochromatin establishment[144]. However, in interphase nuclei, larval neuroblasts and S2 cells, PIAS does not co-localise with heterochromatin and is found near the nuclear periphery, in intranuclear foci, proximal to telomeres and at euchromatin[57].

PIAS proteins are highly conserved across phyla, with homologs in yeasts, worms and plants (Figure 2–2)[145]. In *Drosophila melanogaster* PIAS isoforms share between 36-44% identity with vertebrate PIAS proteins, 29% identity to Gei-17 isoforms in *C. elegans* and 20% identity to Siz proteins in *Arabidopsis thaliana* and *Saccharomyces cerevisiae* (Figure 2–2). In vertebrates, there are 4 PIAS paralogs: PIAS1, PIAS2 (PIAS x), PIAS 3 and PIAS 4 (PIASγ) resulting from a gene duplication event[145].

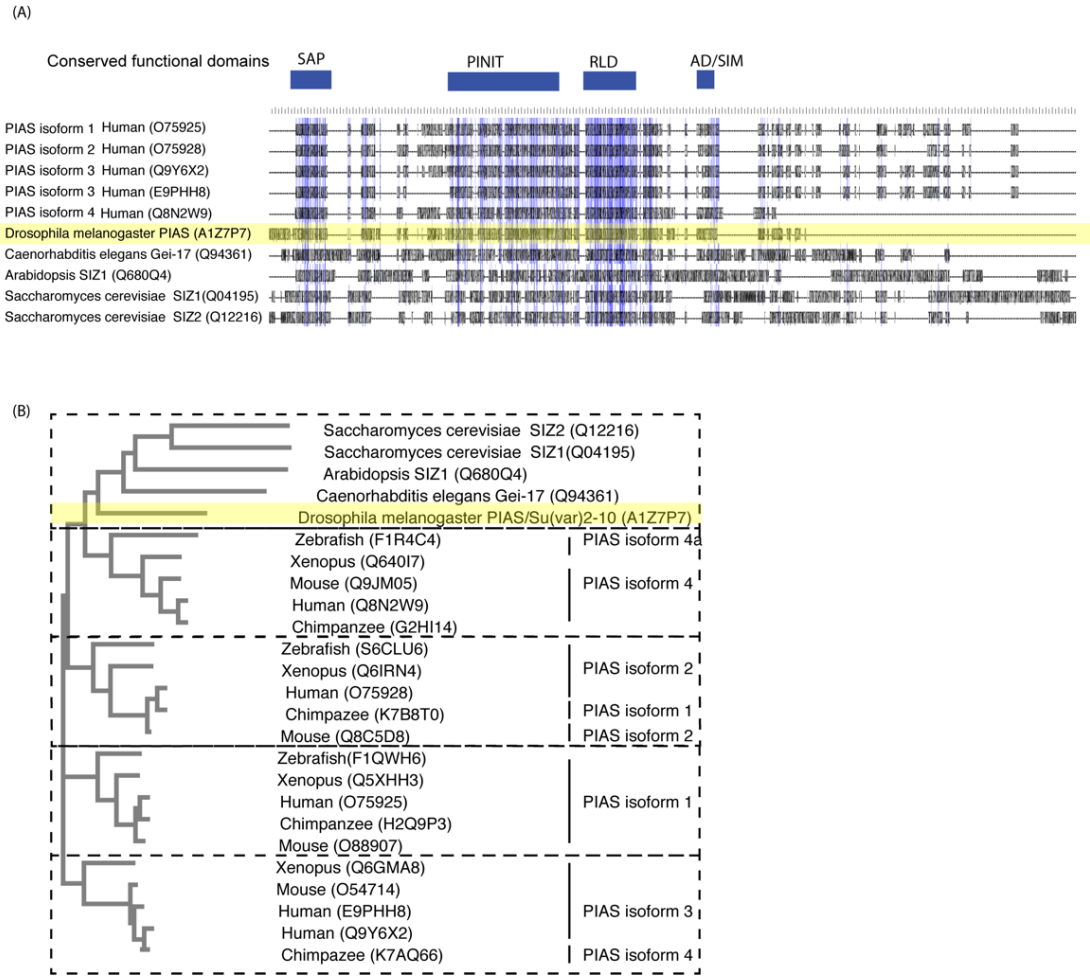


Figure 2–2 PIAS is conserved

(A) Multiple sequence alignment (MSA) of PIAS proteins from different organisms. The protein architecture of PIAS family members is conserved. The position of four conserved functional domains are shown. Conserved amino acids are shown as purple bars. Black lines show the aligned amino acid sequence. Gaps in the sequence occur as a result of alignment.

(B) Phylogenetic tree showing the evolutionary relationships among organisms. Highlighted in yellow is the *Drosophila* PIAS protein where it splits from vertebrates. For *Drosophila melanogaster* PIAS isoform I was used as it is the longest isoform.

The MSA and phylogenetic tree were generated by ClustalOmega, exported into Jalview and exported as SVGs. Protein sequences were downloaded from UniProt in January 2019.

2.2.3 The Structure of PIAS Proteins

PIAS proteins have four conserved functional domains: scaffold attachment factor-A/B/acinus/PIAS (SAP); PINIT; Ring-like domain (RLD); and SUMO interacting motif (SIM). The SAP domain binds AT-rich DNA[146] and associates with the nuclear matrix[147]. The PINIT domain is involved in nuclear retention[148] and substrate recognition by Ubc9[149]. The RLD binds Ubc9 to aid SUMOylation[149,150] of target proteins. The SIM domain can interact with SUMO and SUMOylated proteins[151].

2.2.4 *Drosophila* PIAS Isoforms

14 isoforms of dPIAS have been identified. Different isoforms have different N- and C- termini generated through alternative splicing. A schematic of the mRNA and protein products derived from the *Su(var)2-10* locus is depicted in Figure 2–3. The C-terminal region has intrinsically disordered regions with polar amino acids that allow post-translational modifications and provide platforms for interactions. The isoforms are differentially expressed at different stages in development and in specific tissues[144]. The role of specific isoforms is currently unknown. Exons 2-8 are present in all dPIAS isoforms and contain the functional domains: SAP, PINIT, RLD and AD/SIM. This is referred to as PIAS-CORE. PIAS-CORE is partially sufficient to rescue lethality of *Su(var)2-10*^{1}*/Su(var)2-10*^{2} trans-heterozygous *Drosophila* mutants, but results in infertility[144].

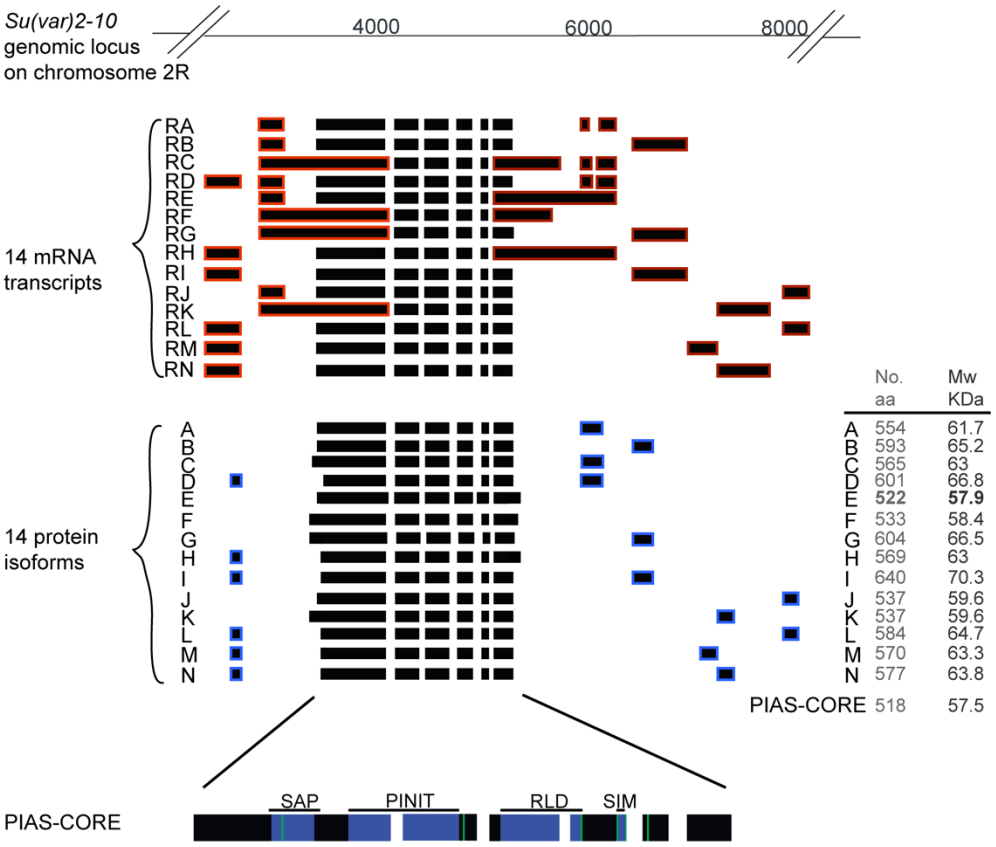


Figure 2–3 Graphic of the *Su(var)2-10* gene locus and its mRNA and protein products

mRNA is transcribed from the *Su(var)2-10* gene locus and is processed, removing intervening regions of DNA (introns), to leave alternatively spliced exons. Black boxes depict mRNA coding regions and exons, which are spliced together in different combinations, giving rise to 14 different transcripts (RA-RN). The protein coding regions (CDS) give rise to 14 PIAS isoforms (A-N). Each dPIAS isoform includes the core region, termed PIAS-CORE, which contains the functional domains: scaffold attachment factor-A/B/acinus/PIAS (SAP); PINIT; Ring-like domain (RLD); and the SUMO interacting motif (SIM). The graphic was generated by aligning all of the PIAS coding sequences (CDS) and mRNA transcripts downloaded from FlyBase (http://flybase.org) in January 2019.

2.2.5 *Drosophila* PIAS Mutant Phenotype

Several *Drosophila* lines generated by Reuter and Wolff[131] have aided the study of the function of dPIAS. The *Su(var)2-10^1 Drosophila* line contains an EMS-induced point mutation in a conserved region of the *Su(var)2-10* locus in the RLD. In *Saccharomyces cerevisiae*, mutation of the equivalent residue in the PIAS homologue Siz abolishes Siz ligase activity[149]. A second *Drosophila* line, *Su(var)2-10^2* has a point mutation changing a tryptophan-to-STOP in amino acid 260, inside the PINIT domain, that results in a truncated version of dPIAS. Trans-heterozygous *Su(var)2-10^1/Su(var)2-10^2* mutants die as third instar larvae displaying melanotic tumours[152], which are also observed in the absence of components of the SUMOylation machinery[153]. These tumours display haematopoietic progenitor states[153] and SUMOylation is disrupted in several types of haematopoietic malignancies[154]. All supporting a hypothesis that PIAS' SUMOylation function is important. Trans-heterozygous *Su(var)2-10^1/Su(var)2-10^2* mutant *Drosophila* also have aberrantly condensed metaphase chromosomes and telomere defects, with telomeres not associating with the nuclear-lamina[141] indicating a failure in chromatin organisation.

2.2.6 The Function of PIAS

PIAS1 and PIAS3 were originally identified in mice as binding partners of the transcription factors STAT1 and STAT3 respectively. PIAS proteins inhibit STAT binding to DNA, preventing transcription of STAT target genes[155-157]. PIAS proteins have SUMO E3 ligase activity, aiding SUMO conjugation to target proteins (discussed in section 2.3.2)[150]. In mice only PIAS1 knockout is embryonic lethal[158]. PIAS1 is critical for cardiac development during embryogenesis, possibly through SUMOylation of the transcription factor GATA4[159-162] which is essential, although further research is required to elucidate the exact mechanisms involved. Deletion of all four PIAS genes has not been attempted.

Drosophila PIAS plays a role in transcription[141], the mitotic cell cycle[141], chromatin structure[141] and has recently been linked to the piRNA pathway[163], after it was identified in 2013[164] in a genome-wide screen in *Drosophila* ovarian somatic sheet cells as necessary for transposon silencing.

2.2.7 The Localisation of *Drosophila* PIAS

In the syncytial embryo, where no cell membranes have formed, nuclei divide rapidly and synchronously. PIAS is maternally deposited and during interphase localises to the cytoplasm and intranuclear spots (Figure 2-4A). During synchronous mitosis, PIAS does not localise to metaphase chromosomes but form foci outside the nucleus (Figure 2-4B-C).

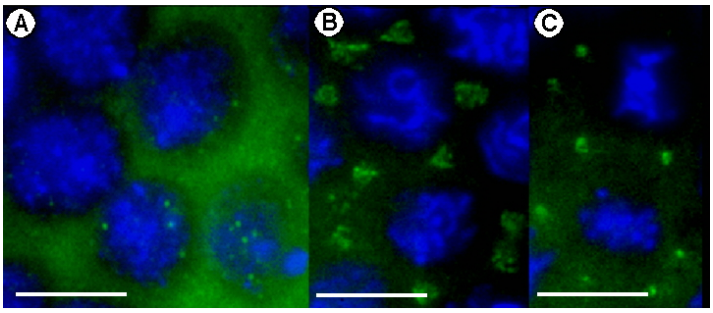


Figure 2–4 Anti-SU(VAR)2-10 staining with exsanguinate sera reveals a dynamic localization pattern for SU(VAR)2-10 proteins in early embryos

In all panels, blue represents DAPI-stained DNA, and green exsanguinate sera staining patterns. **(A)** In interphase/early prophase, spots of staining are found in the nuclear interior. During prometaphase **(B)**, the staining appears to collect into discrete structures, the nature of which is unknown. **(C)** By metaphase, the concentrated staining dissipates. Bar, 5μm. Data from Kumar Hari's thesis, page 81 Figure 4-4[143].

Polytene chromosomes are traditionally used for high-resolution mapping of chromatin proteins. They allow visualisation of inter band patterns of euchromatin and heterochromatin. Kumar Hari showed that PIAS localises to euchromatic bands (Figure 2–5A), telomeric regions (Figure 2–5B) and the heterochromatic chromocenter (Figure 2–5C) [143]. At telomeric regions PIAS is proximal to HP1a and the two proteins do not co-localise.

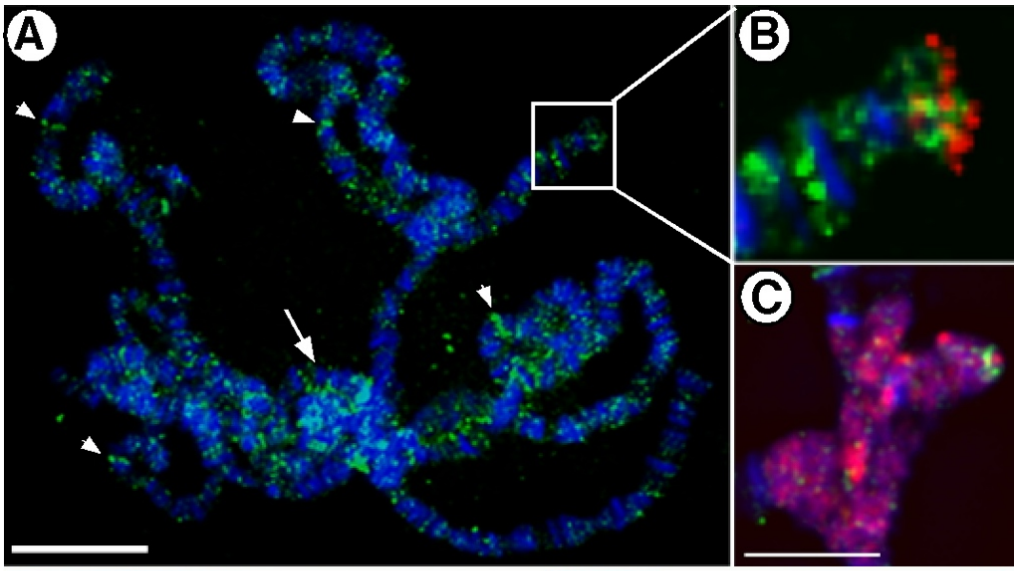


Figure 2–5 SU(VAR)2-10 is generally associated with squashed polytene chromosomes and does not colocalize with HP1 at telomeres

Indirect immunofluorescence showing SU(VAR)2-10 (green) on squashed polytene chromosomes (DAPI stained DNA is blue in all panels). **(A)** SU(VAR)2-10 associates with polytene chromosomes in a general punctate pattern. The chromocenter is labeled (arrow), as are some euchromatic bands (arrowheads), and the tips of chromosome 2L **(B)** and 4 **(C)** Bar in A, 15 µm **(B)** Magnified view of area boxed in A showing that SU(VAR)2-10 (green) does not colocalize with HP1 (red) at the tip of chromosome 2L. **(C)** HP1 (red) densely stains the chromocenter, and chromosome 4, and SU(VAR)2-10 (green) is found near the tip of chromosome 4. Bar, 10µm.

Data from Kumar Hari's thesis, page 83 Figure 4-5[143].

Subsequently PIAS's genomic binding profile was identified, as one of 35 proteins profiled by DamID, to determine different types of chromatin in *Drosophila* Kc167 cells[57]. This found that PIAS is bound to actively transcribed euchromatin in a similar localisation pattern to the insulator protein BEAF-32[57]. PIAS localisation at heterochromatin/euchromatin boundary elements supports a role in chromatin organisation.

2.3 Small Ubiquitin-like Modifiers (SUMO)

2.3.1 Ubiquitin-Like Modifiers (UBLs)

Ubiquitin and ubiquitin-like modifiers (UBLs), including the small ubiquitin-like modifier (SUMO), are a conserved family of proteins with similar three-dimensional structures[165]. SUMO was first identified in 1995, as a suppressor of MIF Two Nr. 3 (SMT3) in *S. cerevisiae*[167,168]. An excess of SMT3 was able to compensate for the phenotypes associated with knockout of MIF Two/Centromeric protein- C (Cenp-C) in *S. cerevisiae*. Cenp-C has since been shown to be SUMOylated[169,170], but the functional relevance remains unknown. UBLs are conjugated to proteins to regulate function. Whereas ubiquitin targets proteins for degradation, SUMO can be reversibly added to substrate proteins to play multiple roles (discussed below)[166].

2.3.2 SUMOylation Pathway

SUMO has to undergo modification itself before it can be conjugated to a target substrate. Since there is high conservation between the machinery of the SUMOylation pathway in *Drosophila* and mammals[171] (e.g Human SUMO-2 shares 68% identity with SMT3, Figure 2–6A), here, I describe the SUMOylation pathway in mammals with the *Drosophila* names in brackets. The process is shown in Figure 2–6B. Firstly, SUMO (SMT3) is made competent for conjugation by cleavage of a C-terminal dipeptide by a SUMO-specific protease (SENP) (ULP) thereby exposing a C-terminal di-Glycine motif (GG). Conjugation competent SUMO (SMT3) is then activated by the E1 heterodimer SAE1/SAE2 (Aos1/Uba2) in an ATP-dependent manner. Next, the E2 conjugating enzyme UBC9 (lesswright) binds SUMO (SMT3) via a cysteine residue and catalyses a transesterification reaction resulting in a highly reactive species. UBC9 (lesswright) catalyses the covalent attachment of SUMO (SMT3) to target proteins at lysine residues. Although not essential for the addition of SUMO (SMT3) to substrates, E3 ligases facilitate conjugation, by providing substrate specificity to UBC9[172], by bringing UBC9 and the substrate within close proximity, which enhances SUMOylation[173]. Substrates can be modified at multiple sites by SUMO (multi-SUMOylation). SUMO can form chains with multiple moieties on one site (poly-SUMOylation) and can form chains with alternative PTMs. This allows the build-up of more complex interaction networks within the cell.

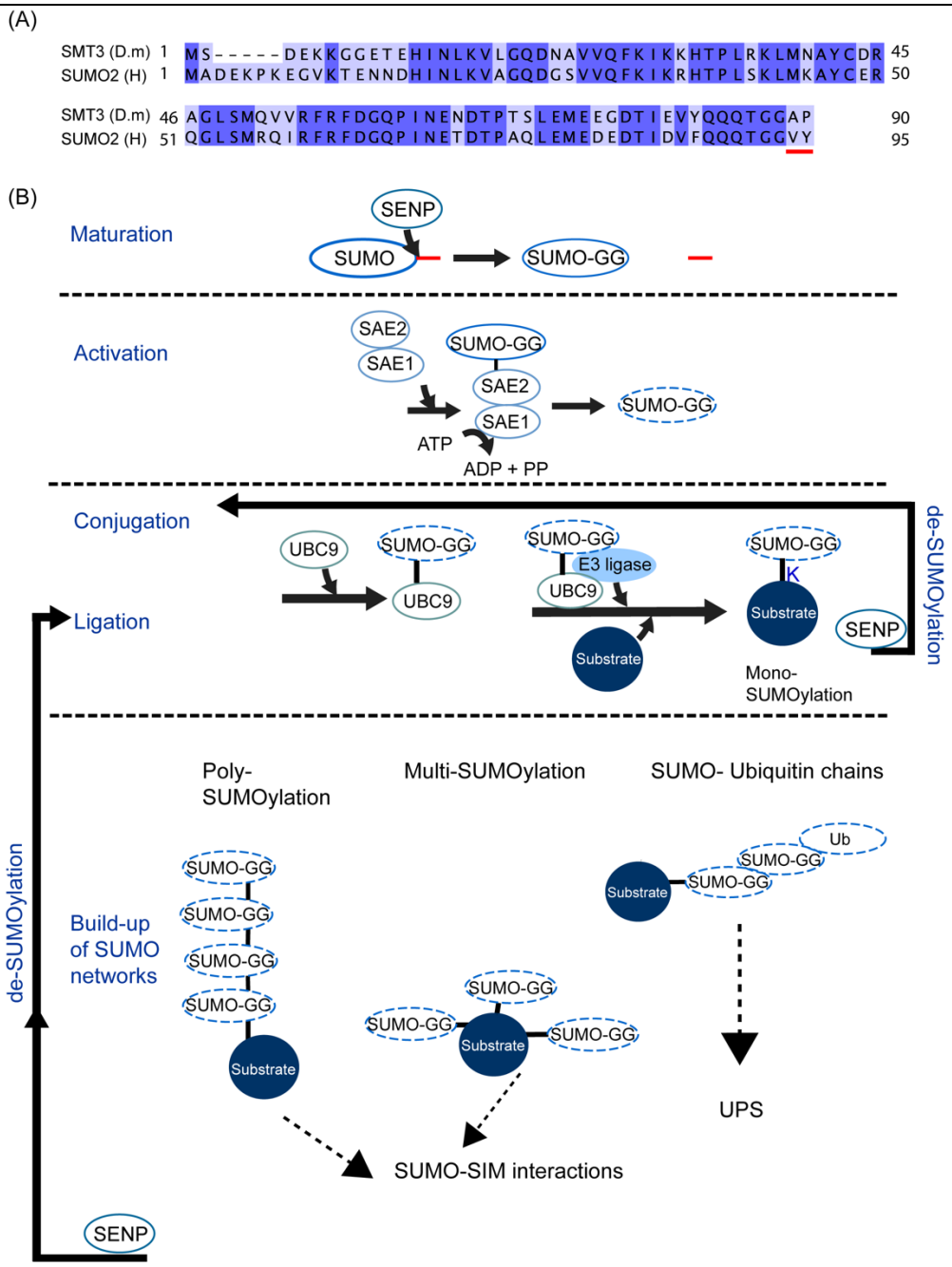


Figure 2–6 The SUMOylation cycle

(A) The amino acid sequence of *Drosophila melanogaster* (D.m) SMT3 aligned to human (H) SUMO2. **(B)** The process of SUMOylation and poly- multi- and SUMO-ubiquitin chain formation (for details of the process see text). SUMO modification on substrate proteins is dynamic and reversible.

The two amino acids at the C-terminus of SUMO/SMT3 that are cleaved to activate it are underlined in red.

2.3.3 SUMO Turnover and use of N-ethylmaleimide (NEM) to Inhibit deSUMOylation

UBLs have limited pools available for conjugation to substrate proteins. SENPs can cleave and release SUMO from substrates to enable conjugation to different subsets of substrates. Under non-stress conditions, for many proteins, only a small fraction of substrate is SUMOylated at any one time, suggesting SUMOylation is controlled dynamically to modulate protein function.

UBLs are constantly turned over *in vivo* and this continues *in vitro* following cell lysis; therefore modified forms of a protein often go undetected when analysed by western blot. The small-molecule inhibitor NEM, is commonly used in lysis buffers to inhibit protein deSUMOylation. NEM is a cysteine peptidase inhibitor which acts by covalently modifying the thiol groups of cysteine residues in their active site (Figure 2–7). SENPs/ULPs are cysteine peptidase which are inhibited by NEM, thus preventing their enzymatic activity to remove UBLs.

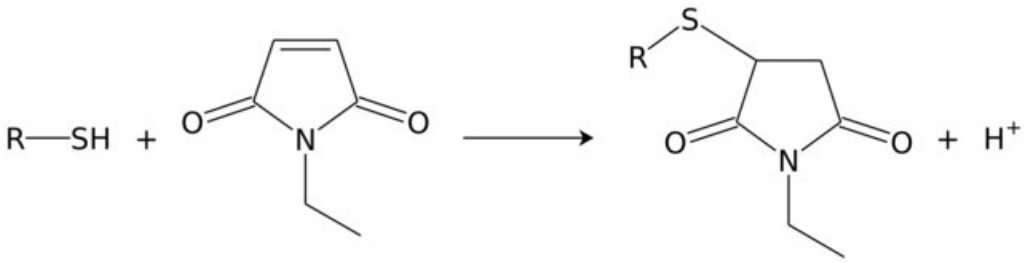


Figure 2–7 The general reaction scheme for N-ethylmaleimide on biological thiols
Image taken from Sasidharan *et al.,* 2012[174].

2.3.4 SUMO Conjugation Sites

The consensus sequence for SUMO conjugation on a substrate protein is the lysine residue contained within the amino acid sequence ΨKxE/D, with Ψ being a hydrophobic residue and x being any amino acid[175]. However, it is estimated that 40% of SUMOylation does not happen at this site[173], with the E3 ligase aiding SUMOylation at non-consensus sites[176]. Tools have been developed to predict SUMOylation sites on proteins by comparison to consensus sequences for SUMO conjugation and experimental data mined from databases[177,178].

2.3.5 Interactions of the SUMO-interacting motif (SIM) with SUMO

In addition to forming covalent interactions, SUMO forms non-covalent interactions with proteins[179]. The best characterised interaction of SUMO with a substrate protein is through a SUMO interaction motif (SIM)[180]. SIMs are characterised by a hydrophobic core of three to four aliphatic amino acids (isoleucine, leucine or valine) flanked by a stretch of acidic amino acids. This promotes an electrostatic SUMO-SIM interaction[181]. Some SIMs are phosphorylation-dependent, with phosphorylation preventing or enhancing binding to SUMO[182].

2.3.6 Biophysical Effects of SUMOylation

When conjugated to a substrate, SUMO alters the biological activity of that substrate by changing the substrate's interaction surfaces. For example, for HP1a, which is a 26kDa protein, the conjugation of a 10kDa SMT3 protein increases its size substantially. Moreover, SUMO chains increase the mass and surface area of small targets. These modifications either prevent or enhance interactions between proteins or between proteins and DNA or RNA. The four main mechanisms through which SUMO can alter a protein's function are shown in Figure 2–8 and described below:

(A) Direct binding of SUMO to a single protein can facilitate interactions, allowing binding of SIM-containing proteins to SUMOylated substrates **(Ai)**; or blocking interactions by preventing recognition of interaction surfaces **(Aii)**; or both can modulate DNA-binding activity.

(B) SUMOylation can modulate or crosstalk with other post-translational modifications (PTMs). Some PTMs target the same amino acid residue on the same protein leading to competition for modification of that site. For example, lysine residues can be modified by: SUMOylation, methylation, phosphorylation and ubiquitination. Therefore SUMOylation can influence the activity of a protein by shifting the balance between its other modifications. Other PTMs are complementary or antagonistic, for example, recognition of SUMO sites within some substrates requires prior phosphorylation of nearby residues[183].

(C) SUMOylation can either recruit ubiquitin (Ub) or prevent it from being added. Ubiquitin-SUMO chain interaction with SUMO-Targeted Ubiquitin Ligases (STubls) results in an altered stability of the protein, resulting in protein degradation by the ubiquitin-proteasome system (UPS)[184].

(D) SUMOylation can have an indirect effect through interactions mediated through SUMO-SIM domains. This can be achieved by sequestering proteins in distinct subcellular structures, altering subcellular localisation **(Di),** forming complex multi-protein complexes **(Dii)** and recruiting protein complexes **(Diii)**.

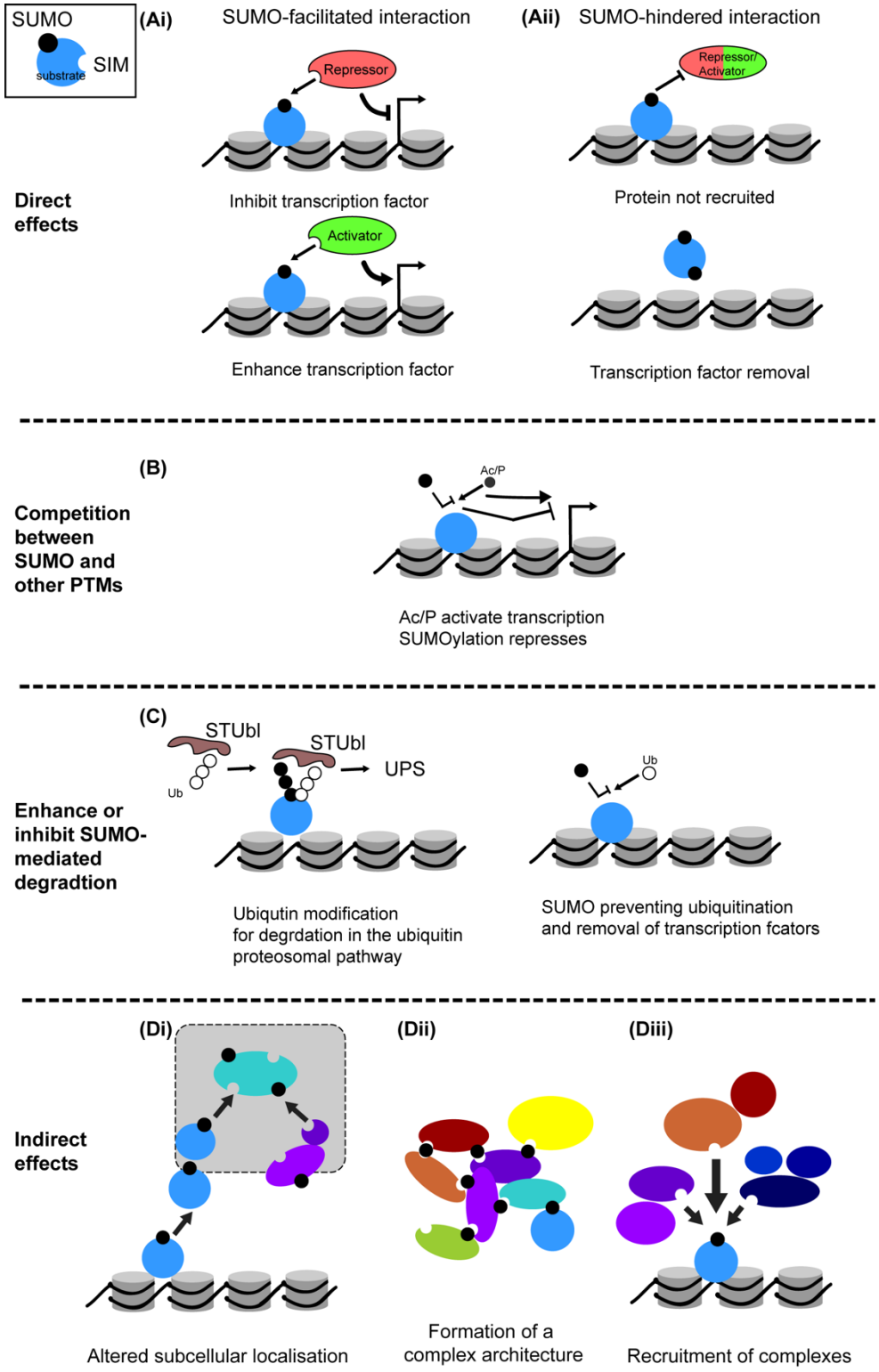


Figure 2–8 SUMOylated proteins and/or proteins with SIM domains act through different mechanisms to modulate protein function

Adapted from Garcia-Dominguez Mario, 2013[185] and Hannoun, 2010[186].

2.3.7 SUMO Plays Essential Roles at a Cell/Organisms Level

Vertebrates have three SUMO paralogs, SUMO-1, SUMO-2 and SUMO-3. SUMO-2 and SUMO-3 share 95% identity but only share 50% identity with SUMO-1[187]. Loss of SUMO2 is embryonic lethal whilst SUMO3 KO mice are viable[188] and SUMO2/3 compensates for loss of SUMO1[189]. In mice, Ubc9-deficient embryos die early post implantation[190].

Invertebrates and yeasts have a single gene encoding SUMO, *Smt3*. *Drosophila* SMT3 is 68% identical to SUMO2 (Figure 2–6). Disregulation of the SUMOylation pathway by reducing lesswright[191], SMT3[171,192], the E1 activating enzyme[192], or an E3 SUMO ligase[152] all yield developmental defects (reviewed in Smith et al. 2012[193]). Conjugation of SUMO to target proteins is essential in all model organisms tested[194] except for *S. pombe,* where *smt3* deletion alters the pattern of histone modifications in heterochromatic regions and causes defects in heterochromatic silencing[195].

SUMOylation is involved in: cell cycle regulation; maintenance of genome stability by influencing DNA recombination and repair; transcriptional regulation; mRNA metabolism[196,197], specifically splicing of pre-mRNAs[198]; nuclear transport; and cell signalling. SUMOylation of substrates increases during stress response, viral infection and cellular senescence. SUMO modification of a specific subset of proteins also acts to stabilise cell fate decisions.

2.3.8 Use of Proteomics to Identify SUMO targets

Approximately 60% of all nuclear proteins have been identified as SUMO substrates with SUMO conjugated on average to 18% of all lysine residues in SUMO target proteins. In human cells over 30,000 proteins were found to be SUMOylated when assayed by mass spectrometry[199]. In a recent proteomics study involving SUMO-2/3 immunoprecipitation to enrich for endogenously SUMOylated proteins from HEK, HeLa, and U2OS cells under heat shock and proteasome inhibition conditions, 67% of the proteins identified were previously reported SUMO targets[200], showing little overlap between experiments performed in different conditions. SUMOylated proteins have been identified in proteomics studies from human[169,201,202], mouse[200,203-205], *C.elegans*[194,206], *Drosophila*[171,207] and yeasts[208]. The effect of protein SUMOylation has been addressed by mutating SUMO sites and assaying the loss of SUMO conjugation phenotype.

Hendriks *et al.,* 2018[200] identified differences in SUMOylation substrates in different cell types and tissues of wild-type mice. The lowest levels of SUMO conjugation were found in the heart and brain and the highest levels in the liver, testis, kidney and spleen. In the brain 78% of SUMO2/3 exists in its unbound state, roughly 10% was found to be conjugated to E1 and E2 enzymes, or integrated into SUMO-chains and the remaining 10% was conjugated to substrate proteins[200]. Henricks *et al.,* 2018[200] suggest that the SUMOylation system exists in a primed state, able to respond to neurological stress, for neuroprotection, as in the case for mice with brain ischemia which have increased levels of conjugated SUMO[209]. In HEK293 cells, they found 93% of SUMO2/3 conjugated to target substrates whereas on average, only 52% of SUMO2/3 was conjugated across all organs. Global remodelling of the SUMO proteome in response to stresses, such as ischemia, salt stress, heat stress or oxidative stress should also be modelled *in vivo* in the targets to be studied to understand the relevance of that SUMOylation; otherwise it is possible that SUMO mediated regulation is possible but does not occurr in a physiological system.

2.3.8.1 Strategy to Identify SUMO targets using Proteomics

Although the role of SUMOylation has in part been studied by detecting which proteins are SUMOylated, proteomics experiments typically relied on the overexpression of transgenes containing modified versions of SUMO[210]. To allow recognition and isolation of SUMO-associated complexes, a standard approach is to add a 6-HIS tag at the N-terminus and modify the C-terminal sequence, exchanging threonine to lysine (T86K), to insert a site for tryptic digest[210]. Modified SUMO immunoprecipitated from cell lysate using Ni^{2+}-NTA affinity resin enabled detection of SUMO substrates by mass spectrometry, detecting a -KGG motif at sites of SUMO conjugation on substrate proteins (Figure 2–9). Algorithms have been developed to detect SUMO binding motifs on peptides. The requirements for detection of SUMOylation sites using the Andromeda peptide search engine are in Appendix 5.

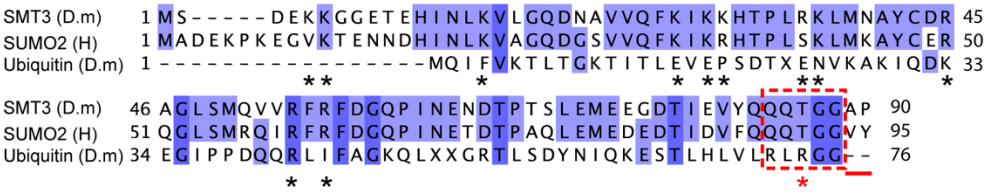


Figure 2–9 Multiple sequence alignment of the amino acid sequences of *Drosophila* SMT3 (O97102), Human SUMO2 (P61956) and *Drosophila* ubiquitin (Q7JPZ2)

Multiple sequence alignment of *Drosophila* (D.m) SMT3 (O97102), Human (H) SUMO2 (P61956) and *Drosophila* (D.m) ubiquitin (Q7JPZ2). Red line marks the C-terminal di-peptide that is cleaved to activate SMT3/SUMO and the box marks the exposed C-terminal regions of these proteins. Lysine (K) and arginine (R) residues are marked by *. SUMO modification can be added at K residues. K and R sites are recognised by trypsin for cleaving the proteins into peptides.

2.3.9 SUMOylation of Transcription Factors

Transcription factors (TFs) are among the most frequently detected targets of SUMOylation, with proteomics experiments estimating that 50% of RNA Pol II TFs are SUMOylated and so are subject to regulation by SUMO[169]. The effects of TF SUMOylation have been studied for about 200 individual TFs (reviewed in Rosonina 2017[211]). With some exceptions, SUMO is generally associated with reduced transcription by a variety of mechanisms (Figure 2–8). A model has been proposed whereby SUMOylation of TFs enhances their DNA binding and protein-protein interactions to control binding site selection and retention[212]. Interesting examples are the transcription factors involved in cellular reprogramming, where hypoSUMOylation enhances pluripotency while increased SUMOylation results in lineage differentiation[213].

Drosophila STAT (STAT92E) acts as a general transcription factor binding zygotic genes at the onset of zygotic genome activation[214]. STAT92E SUMOylation by PIAS[144] inhibits STAT92E binding to promoters, thus inhibiting the JAK/STAT signalling pathway[215].

2.3.10 SUMOylation in RNA processing

SUMO regulates mechanisms involved in RNA processing and stability through modifying the actions of the protein regulators of these pathways. SUMOylated polyadenylation factors disrupt pre-mRNA 3′ processing[216,217] and SUMOylated heterogeneous nuclear ribonucleoproteins (hnRNP) proteins have altered RNA binding properties[218]. SUMO-1 modification of methyltransferase like 3 (METTL3), a component of the RNA N6-adenosine-methyltransferase complex responsible for N6- methyladenosine (m6A), results in repression of its activity[219].

PIAS1, TFs and Tafs co-purify with the mammalian spliceosome[220], suggesting that SUMO conjugation to factors involved in mRNA processing aided by PIAS may influence pre-mRNA processing. PIAS1 is also found in nuclear speckles[221] which are sites in the nucleus enriched for pre-mRNA splicing factors[222].

2.3.11 SUMO "Spray"

SUMO can be conjugated on multiple proteins at multiple sites, resulting in the term "SUMO spray"[223], where SUMO restricted to a local area attaches to substrates to "glue" proteins together and stabilise complexes[224]. Alternatively, "SUMO spray" within a local environment may be ligated to target proteins by E3 ligases, models have been proposed for group SUMOylation[212,223]. Most proteins have multiple SUMOylation sites which leads to redundancy and has created problems in studying and interpreting phenotypes of SUMOylation mutants[225].

2.3.12 SUMOylation in 3D Genome Organisation

The Corces lab has studied the epigenetic regulator protein Modifier of mdg4 (Mod(mdg4)) in the context of its interaction with DNA-binding insulator proteins to regulate transcription of the retrotransposon MDG4 (*gypsy*) in *Drosophila melanogaster*. SUMOylation of Mod(mdg4) inhibits insulator complex formation which is responsible for the formation of chromatin loops[226,227] which determine nuclear architecture. Topoisomerase I-binding arginine/serine rich (Topors) protein physically interacts with components of *gypsy* chromatin insulators and has both E3 SUMO ligase[228] and E3 ubiquitin ligase[229] activity. Ubiquitin modified Mod(mdg4) is required for lamina associations of the *gypsy* insulator and negative regulation of SUMOylation of insulator components. However Topors is unnecessary for *gypsy* chromatin insulator function[230], suggesting functional redundancy amongst mechanisms of SUMOylation and ubiquitination.

2.3.13 SUMOylation as an Initiator in Heterochromatin Formation in Mammals?

SUMOylation of chromatin has been associated with reduced transcription[231,232]. Proteomic studies identified lysine 12 on histone 4 (H4K12) as a specific site of SUMOylation *in vivo*[233]. Studies of this, via chemical SUMO ligation onto histone 4, forces an open chromatin structure[234,235] which can recruit repressive complexes. A SIM in co-repressor of neuron-restrictive silencer element (NRSE)/REI silencing transcription factor (CoREST) binds SUMOylated histone H4 and stimulates lysine-specific histone demethylase1(LSD1)-mediated demethylation of histones[234,235]. CoREST interacts with a PIAS protein (PIASγ) resulting in SUMOylation of CoREST and thereby stabilisation of the CoREST co-repressor complex[236]. In *Drosophila*, the *Su(var)3-3* gene encodes a homolog of LSD1 which acts as an H3K4 demethylase in early embryonic development. This restricts H3K4me2 at heterochromatin through interaction with SU(VAR)3-9, HP1a and RPD3 (HDAC1 in mammals) which removes acetyl marks from N-terminal histone tails[44].

PIAS proteins play a role in mediating SUMO conjugation to chromatin-associated proteins in mammalian cells. One example is SUMOylation of HDAC1[237,238] which aids[239] or inhibits[240] its binding to interaction partners. A second example is SUMOylation of methyl-binding-domain protein 1 (MBD1), which modulates its interaction with CAF-1 and SETDB1 (KMT), controlling cross-talk between DNA and histone methylation. Interestingly, SUMO-1 modified MBD1 cannot form a repressive complex at the promoter of one of its target genes, causing aberrant gene expression[241]. However, the opposite effect was observed in an *in vitro* system with SUMO2/3 modified MBD1 which was able to interact with MCAF-1 and recruit SETDB1[242] to form a repressive complex. The difference in experimental set-up was suggested for the differences in observations. A third example involves AGO2 SUMOylation by PIAS3. This may regulate AGO2's role in RNA interference, with SUMO and phosphorylation possibly cross-talking under hypoxic conditions, to mediate AGO2 binding to Dicer to control gene expression in eukaryotic cells[243].

In mouse cells, HP1α SUMOylation is proposed as a “seeding” step for *de novo* targeting of HP1α to pericentric heterochromatin[244]. The hinge region of HP1α is required for pericentromeric binding[245]. When SUMOylated in the hinge region, HP1α is targeted to pericentric regions in the absence of Suv39h-dependent H3K9me3[244]. SUMOylated HP1α binds chromatin-associated major satellite RNA, which is transcribed from DNA repeats, acting locally in *cis*. SUMOylation of the hinge domain of HP1α allows a more open conformation that allows HP1α to interact with pericentric RNAs and other heterochromatin proteins, allowing the recruitment of Suv39h1[246] which methylates adjacent H3K9. This leads to the recruitment, stabilisation and spreading of HP1α along chromatin, forming larger heterochromatin domains, which silence DNA.

The Suv39h1/2 (SETDB1/2) histone lysine methyltransferase complex (KMT) has two roles at chromatin. In addition to depositing methyl marks at chromatin by tri-methylation of H3K9, these proteins can also enhance SUMO conjugation to substrate proteins, potentially acting as an novel class of E3 SUMO ligases[99,231,247]. SETDB1 proteins have SUMO conjugation and SIM domains[248]. Different chromatin remodelling proteins that confer the same enzymatic activity act to silence different regions of the genome. Association of complexes required for heterochromatin formation and maintenance are likely to be cell cycle, cell type or cell stage dependent.

Krüppel associated box (KRAB)-associated protein-1 (KAP1), also known as Tripartite motif-containing 28 (TRIM28) or Transcription Intermediary Factor 1 β (TIF1β), is another well-characterised E3 SUMO ligase. This acts as a scaffolding protein for the recruitment of chromatin remodelling and repressor complexes that mediate local heterochromatin formation[249]. KAP1 is regulated by post-translational modifications, with phosphorylation aiding its association with DNA replication factors during S-phase of the cell cycle, and KAP1 forms a complex with Suv39h to re-establish methyl marks at heterochromatin after DNA replication[250].

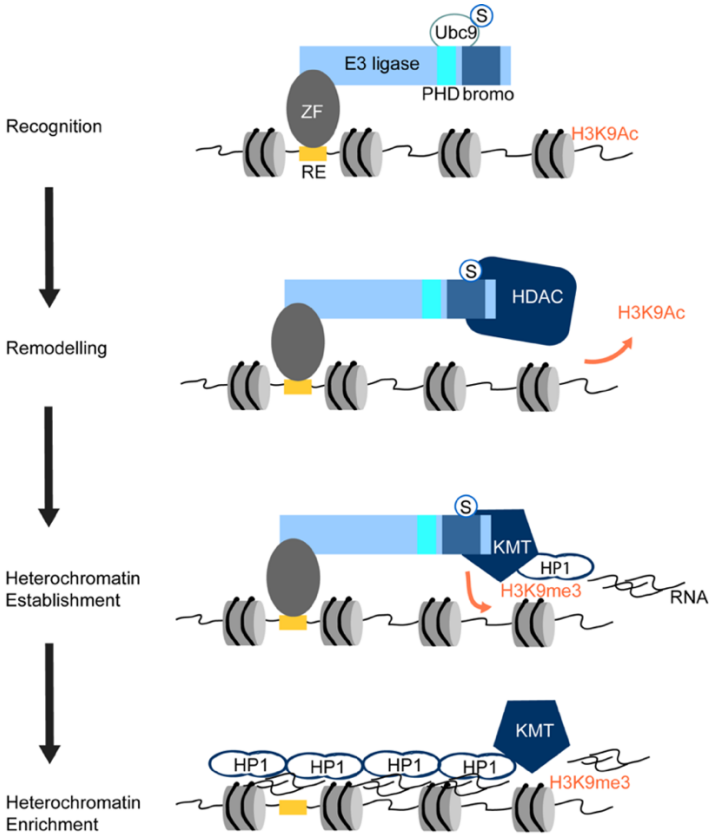


Figure 2–10 Conceptual framework for the establishment and maintenance of constitutive heterochromatin

This model is based on the mammalian KAP1 E3 ligase (E3) binding to recognition elements (RE) through interaction with KRAB zinc finger proteins (ZF). The KAP1 E3 ligase recruits histone deacetylase 1 (HDAC1) and histone lysine methyltransferase (KMT) (SETDB1) for deposition of H3K9me3 and recruitment of heterochromatin protein (HP1). SUMO (S) is required for KAP1 interactions at chromatin. PHD and bromo domains in KAP1 aid chromatin recognition (adapted from Peng 2008[251]) through binding to a combination of post-translational modifications on histone tails.

2.4 The Chromatin Remodeller ATRX

In vertebrates, Alpha thalassemia mental retardation X-linked (ATRX) is a heterochromatic protein encoded on the X-chromosome belonging to the SWI2/SNF2 subfamily of ATP-dependent chromatin remodelling proteins[252,253].

2.4.1 ATRX Domain Structure and Interactions

The ATRX protein has two main functional domains: an ATRX-DNMT3-DNMT3L (ADD), and a SNF alpha-helical domain (Figure 2–11). There is also a third conserved region, the helicase C domain.

The ADD1 domain shares homology with the DNMT3 family of DNA methyltransferases[171]. The ADD domain contains a C2-2 GATA-like zinc finger domain and a plant homeodomain (PHD)[254]. The PHD domain contains an atypical zinc finger which specifically recognises H3K9me3 (but cannot bind H3K4me3 or H3K4me2[255,256]) targeting ATRX to heterochromatin[257,258]. The ADD domain also binds DNA directly at G-quadruplex (G4) DNA structures[259], a feature shared with DMNTs[260]. Binding to both DNA and histones may enhance ADD1 recruitment to specific regions of the genome.

ATRX interacts with HP1[261]. This interaction or ATRX's interaction with methyl-binding protein MECP2[262] may stabilise ATRX at heterochromatin (Figure 2–11).

Figure 2–11 Domain structure of vertebrate ATRX protein

The ADD domain is in the N-terminal region and the SNF/helicase C is in the C-terminal region. Interaction sites for HP1, DAXX and MeCP2 are marked. In purple are the zinc finger (ZF), plant homeodomain (PHD) and an alpha-helical domains within the ADD domain.

2.4.2 ATRX Mutations

In vertebrates ATRX is essential; null mutations in mice are embryonic lethal[263]. Mutation in ATRX can result in a range of mental retardation syndromes[264-268] which vary between individuals and may also include: altered facial features; genital abnormalities; and alpha thalassemia. The mutations cluster over two functional domains; 50% of mutations are located in the N-terminal ADD domain and 30% in the C-terminal SNF (Helicase/ATPase) domain[258].

2.4.3 ATRX Function and Mechanisms at Chromatin

ATRX is required for transcription at satellite repeats, G-rich sequences[269], G4-DNA structures[270], RNA-DNA hybrids (R-loops)[271], pericentric heterochromatin repeats and sub-telomeric heterochromatin repeats[272]. ATRX is also is involved in the formation of heterochromatin structures to silence endogenous retroviruses[273,274], imprinted genes[262] and tandem repeats[275]. Knockdown of ATRX results in mitotic spindle defects, chromosome misalignment, segregation problems[276], an alternative lengthening of telomeres phenotype[277] and impaired transposon silencing[274].

Once recruited to its target sites, ATRX targets the histone chaperone death-domain-associated protein (DAXX) to facilitate incorporation of the histone variant H3.3 into nucleosomes, which may maintain DNA in the B-form conformation to maintain an open chromatin structure[278,279] for replication through these regions (Figure 2–12).

ATRX functions as part of several repressive complexes. In human cells, ATRX recruits Chromobox protein homolog 5 (CBX5) (the human homolog of *Drosophila* HP1a) to telomeric regions, which acts with zinc finger protein ZNF274, TRIM28 and SETDB1 to deposit and maintain H3K9me3 at telomeres at the 3' regions of zinc finger genes[280,281,282]. Mutations in ATRX disrupt the organisation of pericentromeric heterochromatin[46] by altering the pattern of DNA methylation[283]. Absence of ATRX disrupts genome organisation in the brains of neonatal mice[284].

In mammals, telomeric repeat-containing RNAs (TERRA) transcribed from sub-telomeric regions of the genome[285] compete with ATRX for binding to telomeric DNA[286]. TERRA forms R-loops that promote the formation and stabilisation of G4 structures at telomeric repeats. RNA-G4 structures formed by TERRA bind to lysine-specific histone demethylase1 (LSD1), which catalyses the removal of methyl groups from H3K4 and H3K9[287]. This aids telomere recombination[288] to maintain telomere length in the alternative-lengthening of telomeres (ALT) pathway. ALT is activated in many cancers and is critical for cell survival. ATRX supresses the ALT pathway. In the absence of ATRX, TERRA transcription increases and telomeres are derepressed[64]. Therefore ATRX plays a critical role at telomeres.

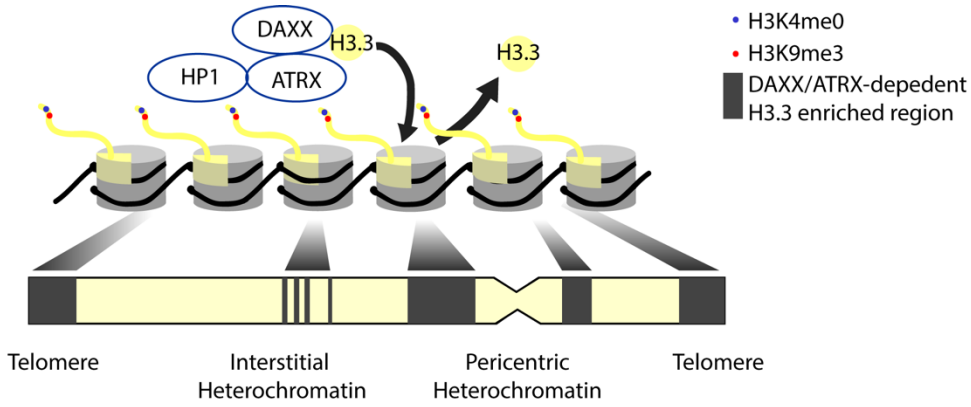


Figure 2–12 Model of ATRX-DAXX dependent targeting of H3.3 to pericentric heterochromatin, interstitial heterochromatin and telomeres

ATRX is recruited to chromatin through its ADD domain which recognises a combination of H3K4me0 and H3K9me3 on histone H3 tails. ATRX interacts with DAXX to aid deposition of the histone variant H3.3 into nucleosomes to maintain an open DNA conformation to allow gene expression.

2.4.4 ATRX Conservation

Functional homologs of ATRX, with a conserved SNF domain structure, have been identified in both vertebrate and non-vertebrate species (Figure 2–13), with a conserved function in the deposition of H3.3 into chromatin in *Drosophila*[289] and *Arabidopsis*[290] and a role in chromatin remodelling in *C.elegans*[291]. Rad54-like homologs of ATRX exist in yeasts[292].

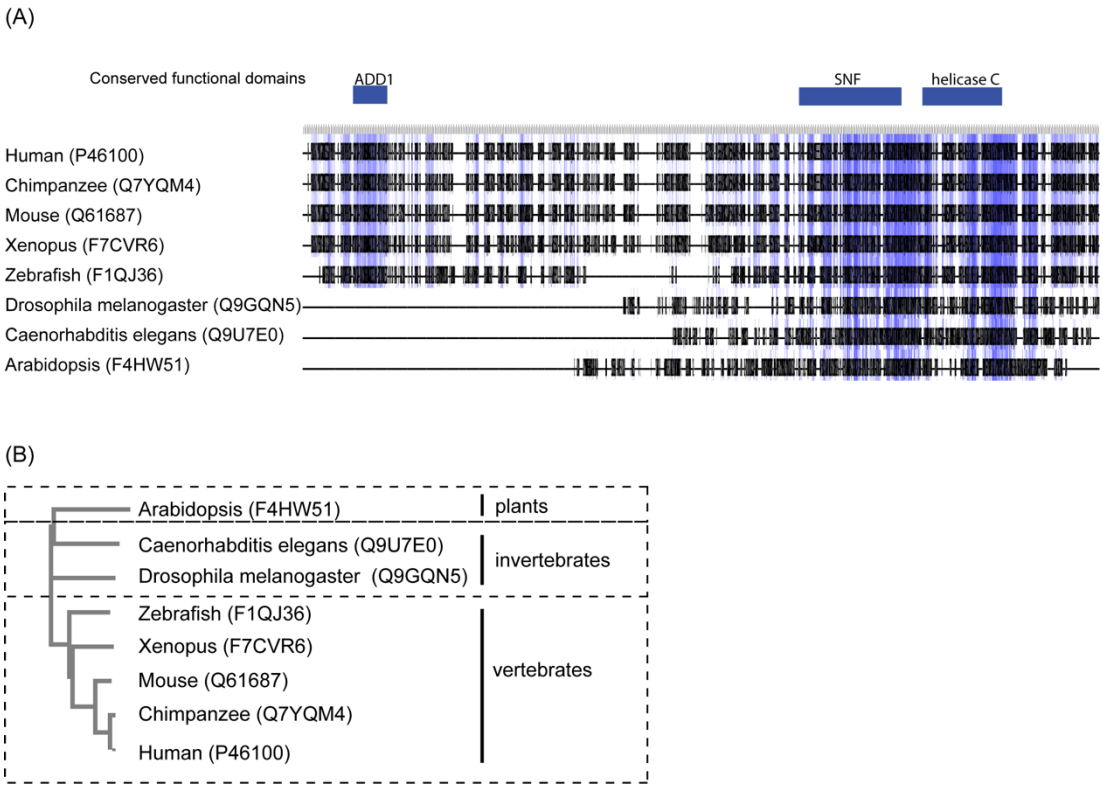


Figure 2–13 Multiple sequence alignment of ATRX proteins from different organisms and the phylogenetic tree derived from it

(A) Multiple sequence alignment (MSA) of ATRX proteins from different organisms. The position of three conserved functional domains is shown. Conserved amino acids are shown as purple bars. Black bars show the aligned amino acid sequence, gaps in sequence occur as a result of alignment. **(B)** Phylogenetic tree showing the evolutionary relationships among organisms. The MSA and phylogenetic tree were generated by ClustalOmega and exported into Jalview to visualise conserved amino acids in purple, then exported as a scalable vector graphic to generate the figure in Adobe Illustrator. Protein sequences were downloaded from UniProt in January 2019.

2.5 XNP, the C-terminal Homolog of ATRX in *Drosophila*

XNP was identified as the functional homolog of ATRX in *Drosophila*, aligning to the C-terminal SNF and helicase C/ATPase domains[139](Figure 2–13). XNP is expressed throughout development as a long and a short isoform[139]. The longer isoform forms a complex with HP1a, is enriched at heterochromatin and facilitates HP1a deposition into pericentric beta-heterochromatin in the X chromosome[293,294]. Knock-out of the long isoform of XNP does not affect the viability of *Drosophila*, whereas loss of both isoforms is semi-lethal (with ATRX pupa emerging at sub-Mendelian ratios) and results in infertility[294].

XNP localises to sites of active transcription in regions of rapid nucleosome turnover[289], at pericentric heterochromatin[139] and telomeres[295]. Interestingly, XNP concentrates at a satellite block adjacent to pericentric heterochromatin on the X-chromosome[289]. This region is highly decondensed and made up from a 4-base pair TAGA DNA repeat. The TAGA repeat is only present in *Drosophila melanogaster*, is adjacent to the chromocenter and is formed of pericentric beta heterochromatin[289].

PIAS[144] and SMT3[144] also localise to the TAGA repeat in *Drosophila melanogaster*. However, ADD1[293] and HP1a[289] do not localise here. HP1a[296], ADD1[296] and XNP[294] all localise to the chromocenter, which is formed from a cluster of pericentric heterochromatin regions and the forth chromosome. It has been suggested that the TAGA repeat is a genome-wide regulator of genome silencing[289]. Tools to visualise repetitive regions in *Drosophila*[297] have shown that pericentric satellite DNA is able to aid DNA packaging[298] and are the subject of much study.

2.6 ADD1, the N-terminal Homolog of ATRX in *Drosophila*

A phylogenetic analysis performed on ATRX homologs suggested that the ATRX gene split in insects[296] resulting in two proteins in *Drosophila*, ADD1 and XNP (Figure 2–14). A protein sharing homology with the N-terminal ADD domain of vertebrate ATRX was discovered[299] sharing the zinc finger and PHD domains (Figure 2–15).

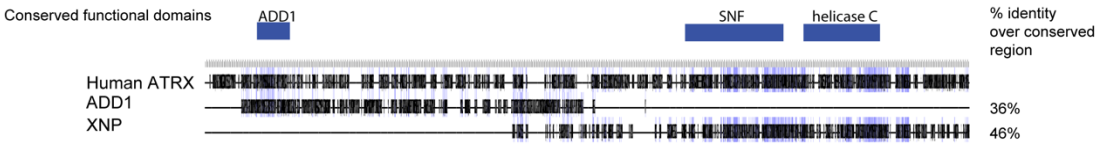


Figure 2–14 Protein multiple sequence alignment of *Drosophila* ADD1 and XNP with human ATRX

The position of conserved functional domains are marked. Also shown is the percent identity over conserved regions of ADD1 to ATRX and XNP to ATRX. MSA was generated by ClustalOmega, exported into Jalview for visualising conserved amino acids as purple bars. Black lines show the aligned amino acid sequence, gaps in sequence occur as a result of alignment. Protein sequences were downloaded from UniProt in January 2019.

Figure 2–15 Protein multiple sequence alignment of the ADD domain of *Drosophila* ADD1 and human ATRX

The conservation of the zinc finger and PHD domains of the ADD domain in ATRX/ADD1 is shown. MSA was generated by ClustalOmega, exported into Jalview for visualising conserved amino acids as purple boxes. Protein sequences A1Z8R2 and P46100 were downloaded from UniProt in January 2019.

Drosophila ADD1 localises to pericentromeric heterochromatin and telomeres, co-localising with HP1a[299]. The *add1* gene encodes three ADD1 protein isoforms generated by alternative splicing[296]. Reduced expression of the longest isoform of ADD1 (ADD1-RA) results in loss of HP1a at telomeric sequences without affecting HP1a localisation to other binding sites[295].

ADD1 interacts with HP1a, repetitive RNAs and the protein BONUS[299]. BONUS is the only known *Drosophila* homolog in the Transcription Intermediary Factor 1 (TIF1) family[300]. In humans TIF1s phosphorylate HP1[301,302] to regulate HP1 binding to DNA[296]. In *Drosophila,* bonus has not been identified as an interactor of HP1[303], but potentially ADD1, BON and HP1a could be part of the same complex *in vivo.*

2.7 ADD1-XNP Interaction

ADD1 and XNP genetically interact. In a genetic screen conducted by Lopez-Falcon *et al.,* 2014[296], the percentage viability of heteroallelic XNP mutants ($atrx^{1}/atrx^{3}$) (49%) is increased in *Drosophila* expressing a hypomorphic allele of ADD1 ($dadd1^{NP1240}/+;atrx^{1}/atrx^{3}$) (72%); however, 12% of these *Drosophila* develop melanotic masses[296]. This suggests an epistatic relationship between ADD1 and XNP.

Lopez-Falcon *et al.,* 2014[296] mapped a direct interaction between amino acid residues 1-620 of ADD1-RA and XNP-RA residues 1-221 (found only in the long isoform of XNP)[296]. This interaction was not found by Alekseyenko *et al.,* 2014[293]. They demonstrate that XNP and ADD1 interact strongly with HP1a but ADD1 pull-down shows no interaction with XNP[293]. The difference in these studies are the methods used to isolate ADD1 associated complexes, suggesting that certain conditions are required to maintain and observe an interaction between ADD1 and XNP.

In *Drosophila* XNP does not constitutively associate with heterochromatin[289]. In contrast, ADD1 localises and interacts with heterochromatin proteins throughout the cell cycle[303]. XNP may localise with ADD1 and HP1 in certain circumstances or at specific chromatin regions.

2.8 SUMOylation of ADD1 and XNP

XNP can be covalently modified by SMT3[171]. In the Heun lab, lysine 181 of XNP was identified as a SUMOylation site (within an N-terminal XNP fragment) in an *in vitro* bacterial SUMOylation assay (presented in the master's thesis of Yvonne Steuernagel[304]). The mammalian ATRX homolog has also been identified as a SUMO substrate in mass spectrometry experiments looking to identify SUMO interactors by immunoprecipitation of SUMO-complexes from human and mouse tissue culture cells[305].

ADD1 was identified in a one-step purification of SUMOylated proteins in *Drosophila* S2 cells in non-denaturing conditions (Heun lab unpublished) where non-covalent modifications are maintained. This raises the question of whether ADD1 is itself SUMOylated or whether ADD1 interacts with SUMOylated targets through SIMs.

2.9 Hypothesis

We hypothesise that PIAS-mediated SUMOylation of chromatin factors is responsible for establishing and/or maintaining heterochromatin.

XNP is SUMOylated in *Drosophila*[171]. It was hypothesised that PIAS may act as the SUMO E3 ligase aiding SUMOylation of XNP. PIAS and XNP co-immunoprecipitate when overexpressed in S2 cells and an intimate relationship between PIAS and XNP was predicted based on their co-localization in S2 cells and *Drosophila* tissues[144]. It was also observed by the previous PhD student in the Heun lab that *S*UMO, PIAS and XNP localise to DAPI dense regions in pre-blastoderm nuclei of *Drosophila* embryos (Figure 2–16)[144]. This localisation coincides with heterochromatinization and thus a PIAS-XNP-SUMO complex may be important for XNP SUMOylation and heterochromatin establishment.

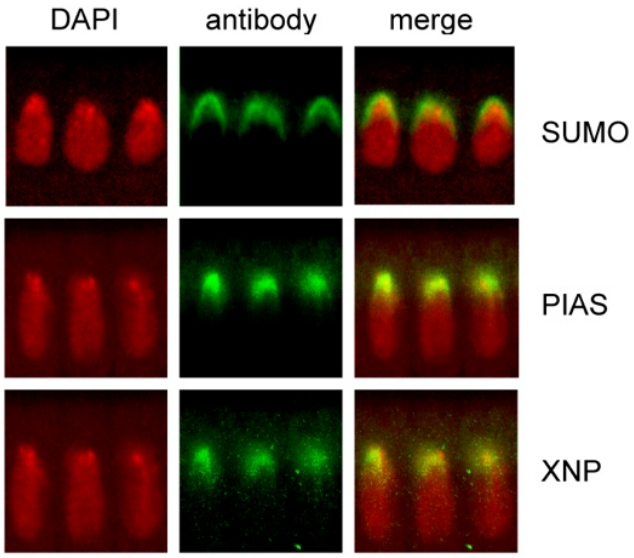


Figure 2–16 Localisation of SUMO, PIAS and XNP to the DAPI dense regions of pre-blastoderm nuclei in *Drosophila* embryos in nuclear cycle 14

The observation was first made by Lorenz Kallenbach and was presented in his thesis in 2011[144]. Images were taken by Lorenz Kallenbach and were compiled by myself for presentation in this thesis.

2.10 Aims of this thesis

PIAS proteins in *Drosophila* can be studied to understand common roles for PIAS in other species. Based upon our hypothesis that PIAS-mediated SUMOylation of chromatin factors is responsible for establishing and/or maintaining heterochromatin, I identified four key aims:

1. to develop tools for identifying PIAS, ADD1 and XNP interactomes at endogenous protein levels
2. to understand whether the previously identified PIAS interaction with XNP occurs under physiological conditions
3. to characterise the interaction between ADD1 and XNP with PIAS
4. to identify other PIAS interactors that may have an effect at heterochromatin.

2.11 Strategies to study PIAS localisation and interaction partners

Previous experiments in the Heun lab, to characterise the localisation and interaction partners of PIAS, were performed by inducing expression of tagged PIAS transgenes from plasmids. In this thesis, use of the *Drosophila* anti-PIAS sheep antibody and CRISPR/Cas9 tagging of PIAS proteins allowed further experiments to be performed, to assess endogenous PIAS's localisation and interaction partners. These strategies are outlined in Figure 2-17.

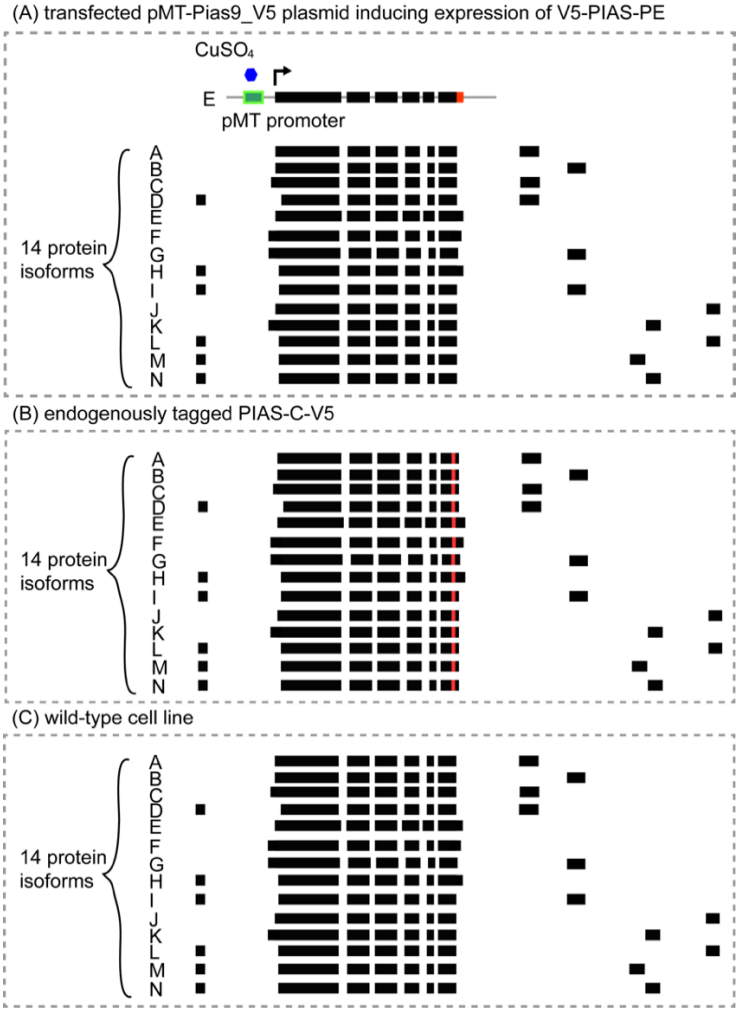


Figure 2-17 Strategies to study PIAS localisation and interaction partners

Drosophila and *Drosophila* S2 cells express 14 isoforms of PIAS. The first strategy involves inducible expression of a V5-PIAS-PE transgene, from a *pMT-Pias9_V5* plasmid stably maintained in an L2-4 S2 cell line, by the addition of copper sulphate **(A**). The second strategy involves tagging all 14 isoforms. Red marks the position of the C-terminal PIAS CORE tag **(B)**. The third strategy uses anti-PIAS antibodies **(C**).

Chapter 3 Methods

3.1 DNA and RNA protocols

3.1.1 gDNA isolation

Qiagen Quick Extract buffer (Qiagen) was used for extraction of gDNA from S2 cells following the manufacturer's instructions.

3.1.2 DNA extraction

For a clean DNA preparation, DNA was prepared from $1x10^6$ S2 cells. S2 cells were resuspended in genomic lysis buffer and were incubated at RT for 10min. Each sample was treated with 5µg/mL RNase, DNase-free (Roche) (0.5µL 500µg/mL stock solution) and incubated at 37°C for 30min, followed by proteinase K (200µg/mL) at 50°C overnight. The following day, proteinase K was heat inactivated at 80°C for 20min. DNA was extracted twice with 50µL Phenol:Chloroform:Isoamyl Alcohol (at ratios 25:24:1, Sigma, 77617) added to the DNA, vortexed then centrifuged for 5min at 10000 x g. The DNA-containing aqueous phase was added to 50µL Chloroform:Isoamyl Alcohol in a fresh tube, vortexed then centrifuged for 4min at 10000 x g. DNA contained in the aqueous phase was then precipitated by adding 50% isopropanol and 150mM NaOAc pH 5.2 and incubating overnight at -20°C. To pellet the precipitated DNA, samples were centrifuged at 18000 x g for 10min at RT. The DNA pellet was then washed in 70% EtOH and air dried at RT for 5min. The DNA was then resuspended in 5µL of ddH_2O.

3.1.3 DNA and RNA quantification

DNA and RNA sample concentration and purity were assessed using a ND-1000 Spectrophotometer (NanoDrop). The spectrum of absorbance and the A_{260} /A_{280} ratio was used as an estimation of DNA/RNA sample purity.

3.1.4 RNA extraction and cDNA synthesis

RNA was extracted from approximately $1x10^7$ S2 cells using the TRIzol method (Sigma). Cells were rinsed once with ice-cold PBS then 2.5mL ice-cold TRIzol was added and incubated for 5min on ice. Cells were resuspended to homogenize the

solution and were transferred to a 15mL falcon and vortexed. RNA was recovered through phase separation by adding 500µL chloroform and vortexing. Samples were incubated at RT for 2min then centrifuged for 15min at 12000 x g at 4°C. The colourless aqueous upper phase was transferred into a fresh 2mL tube.

RNA was precipitated by mixing the aqueous phase with an equal volume of isopropanol, incubated at RT for 10min, then centrifuged for 10min at 12000 x g at 4°C. The RNA pellet was washed with 70% EtOH and resuspended in DEPC-treated water. 15µg RNA was DNase treated with 2µL DNaseI (NEB) in a 50µL reaction for 1hr at 37°C. 5mM EDTA was added and the DNAseI heat inactivated by incubating at 75°C for 5min to stop the reaction. RNA was quantified using the ND-1000 Spectrophotometer (NanoDrop) and LiCl precipitated (-20°C overnight). Samples were aliquoted and stored at -80°C until needed.

First strand cDNA was generated from 1µg of total RNA. Firstly, 1µg RNA, 1µL Random Hexamers (100ng/µL) (Thermo Fisher Scientific) and 1µL dNTPs (10mM), were mixed, and the total volume brought to 13µL with ddH_2O. The samples were denatured at 65°C for 5min and immediately put on ice to prevent re-annealing. Secondly, a mix containing 4µL 5x First-strand buffer, 1µL DTT (from 1M DTT stock), 1µL Superscript III Reverse Transcriptase (RT) (Invitrogen) and 1µL ddH_2O were added to the first tube. Samples were then run on a thermocycler using the following programme: 25°C for 5min, 50°C for 1hr, 70°C for 15min. Samples were diluted 1:20 before PCR. A control sample was made by substituting for the RT with 1µL of ddH_2O (no RT). The concentration and purity of the cDNA were determined using the ND-1000 Spectrophotometer (NanoDrop).

3.1.5 Polymerase Chain Reaction (PCR)

PCR reactions were performed using Phusion® High-Fidelity DNA polymerase (NEB). Elongation time was adjusted on the assumption that processivity is 30s/kb. Extension and annealing temperatures were set according to the predicted melting temperature of the primer pair used. For the template DNA to be amplified, approximately 10ng from plasmid DNA, 1µL gDNA extract or 100ng cDNA was used. Reactions were carried out in 0.2mL thin-walled PCR tubes (STARLAB).

A standard 50µL PCR reaction contained: template DNA; 1x Phusion HF buffer; 0.5µM forward primer; 0.5µM reverse primer; 25µM dATP; 25µM dGTP; 25µM dTTP; 25µM dCTP and 0.5µL Phusion® High-Fidelity DNA polymerase (NEB) brought to 50µL with ddH_2O.

The standard PCR programme was:

95°C for 5min
  95°C for 30s
  55°C* for 30s
  72°C for 30s/kb of product
 [Repeat x 29 cycles]
72°C for 10min
4°C for 5min

*(annealing temperature 5°C below the melting temperature (T_m) of the primers)

3.1.6 Genotyping PCR

PCR reactions were performed using primers designed to amplify gDNA across a target region (for primer sequences see Appendix Table 20). Primers were designed using SnapGene (version 4.3.5) and checked with PrimerBLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/).

3.1.7 Agarose gel electrophoresis

For analysis of PCR products or DNA digestions, DNA was separated according to molecular size by agarose gel electrophoresis. Agarose (0.7%-3% w/v) was dissolved in 1x TAE buffer by heating in a microwave. Once cooled, 0.01% SYBR Safe DNA gel stain (Invitrogen) was added. DNA samples were loaded in 1x DNA Loading Buffer alongside a DNA molecular weight marker for size comparisons. An electric current (90-130V) was applied to the gel for 30-60min. Following electrophoresis, the DNA was visualized on a UV transilluminator Gene Flash Documentation System (Synergene) or on a Safe Imager 2.0 Blue-Light Transilluminator (Invitrogen) for excision of bands from the gel.

3.1.8 Design of CRISPR targeting constructs in S2 cells

500ng of 500bp gene block DNA fragments were designed to act as homology directed repair templates (HDRT) for insertion of a tag into the genomic loci encoding either PIAS, ADD1 and XNP proteins. These HDRTs were designed using the extended gene regions of PIAS, ADD1 and XNP (downloaded from FlyBase on 03-12-2016). Tags, linker regions, restriction sites and a base pair change in the protospacer adjacent motif (PAM) sgRNA recognition sequence were introduced and at least 200bp of homology arms either side of the tag were synthesized by IDT. PIAS and ADD1 HDRT designs used to generate cell lines used as part of this study are outlined in the results section 4.3.12 and appendix 4 respectively.

3.1.9 Cloning of plasmid repair templates for CRISPR targeting

Plasmid cloning was used to generate homology directed repair in *Escherichia coli*. To prepare fragments for cloning, 250ng gene block and 1µg of vector were incubated with restriction enzymes (NEB) (10U/µg DNA to be digested) for 2hr at 37°C. Age1 and Kpn1 were used for cloning of the ADD1 HDRT into pMT_GFP_lacI and KpnI and SacI were used for cloning of PIAS HDRT into pBlue. Digests were separated on 0.7% agarose in 1xTAE buffer at 130V for 30min and the bands of interest were cut from the gel and purified using the GenElute™ Gel Purification Kit (Sigma) following the manufacturer's instructions.

Insert and vector were ligated at ratios 5:1, using 50ng vector DNA in a 20µL reaction, for 2hr at RT with 1µL T4 DNA Ligase in 1x T4 ligase buffer (Thermo Fisher Scientific) following the recommended protocol. A negative reaction, prepared with vector only, was used to assess background self-ligation rate. 1µL of the ligation product was mixed with a 50µL aliquot of electrocompetent 10beta *E. coli* cells and the contents of the tube transferred to a pre-chilled 0.1cm electroporation cuvette. This was electroporated using an Eppendorf Eporator system (Eppendorf) set on P1 1800V. 1 mL of SOC medium was added to the cuvette and the cuvette contents were transferred to sterile tubes which were then shaken for 30min at 200rpm at 37°C in a Thermo Mixer (Eppendorf). 100µL of the *E. coli* culture was then plated on LB plates containing selection antibiotics (see materials section) and grown overnight at 37°C until bacterial colonies could be observed growing on the plate.

Single colonies were picked to inoculate 2mL of LB broth containing selection antibiotic (1000x solutions were prepared as stocks, for concentration of antibiotics see Appendix materials section Table 17) and left overnight in a 37°C shaking incubator (200rpm).

Plasmids were purified from *E. coli* using the GenElute™ Plasmid Miniprep Kit (Sigma) following the manufacturer's instructions and quantified using the ND-1000 Spectrophotometer (NanoDrop). Analytical restriction digests were performed to verify the final product of each construct. These were performed in 10µL reactions using 100ng plasmid DNA and 1U of each restriction enzyme.

3.1.10 Ligation-independent cloning (LIC)

3C LIC vectors, pFL and pFastBac, were obtained from Atlanta Cook for generation of GST and HIS tagged Sf9 protein expression vectors. The plasmid maps are in Appendix 3. 1µg parent vector was digested with 10U ZraI resulting in a linearized plasmid. Once linearized, the plasmid was run on 0.7% agarose and excised for gel-purification using the GenElute™ Gel Purification Kit (Sigma) following the manufacturer's instructions.

For the insert, primers were designed to amplify the gene of interest from cDNA to include overhangs to allow processing for insertion into the 3C LIC vector backbone. The primers span the start and end of the sequence of the gene of interest flanked by LIC sites (Table 1). For PCR, 1µL of first strand cDNA was mixed with 1x Phusion HF buffer, 0.5µM forward primer, 0.5µM reverse primer, 25µM dATP, 25µM dGTP, 25µM dTTP, 25µM dCTP and 0.5µL Phusion® High-Fidelity DNA polymerase (NEB) brought to 50µL with ddH_2O. The PCR was performed in two stages. This included an initial denaturation at 95°C for 3min followed by 13 cycles of denaturation at 95°C for 30s, annealing at 65°C for 30s, and extension at 72°C for 2min. This was followed by 25 cycles of denaturation at 95°C for 30s, annealing at 55°C for 30s, and extension at 72°C for 2min, with a final extension at 72°C for 10min. The PCR products were separated on 0.7% agarose and excised for gel-purification following the manufacturer's instructions.

450ng vector was mixed with 0.45U T4 DNA polymerase LIC-qualified (Novagen 70099-3) in 1x T4 DNA polymerase buffer, 5mM DTT and 2.5mM dTTP in a 30µL reaction. For the insert, 600ng of gel-purified PCR product was mixed with 0.6U T4 DNA polymerase LIC-qualified (Novagen 70099-3) in 1x T4 DNA polymerase buffer, 5mM DTT and 2.5mM dATP in a 20µL reaction. The reactions were left for 30min at RT, then heat inactivated for 20min at 75°C. In the presence of a single nucleotide the T4 DNA polymerase 3'->5'-exonuclease and 5'->3'-polymerase activity is in equilibrium at the site of the first occurrence of this nucleotide, which results in unique single stranded overhangs.

The vector and insert were annealed by mixing 1µL vector:1µL insert and incubating for 10min at RT. Note that this is due to the large size of the insert and deviates from the Cook lab protocol. Addition of 1µL 25mM EDTA was added. After another 10min at RT 2µL of the mix was transformed into electrocompetent 10beta *E. coli* cells for propagation of the protein expression vectors .

Table 1 Ligation independent cloning (LIC) primer sequences

ADD1RA_LIC_FW	ccaggggcccgactcgATGAGTAACAGTGCTCCGGGTAGCGAAAGTGG
ADD1RA_LIC_Rv	cagaccgccaccgactgcttaGCCAACAATCTCAGCCGTGGCGAACTCG
XNP_LIC_FW	ccaggggcccgactcgATGGGAAAGAAAAACCCCAACGCCCGTCACA
XNP_LIC_Rv	cagaccgccaccgactgcttaGTCGATCTCGTATACCTTATCCGGTTCA

Selected colonies were amplified in 2mL cultures and vector DNA was extracted by performing a miniprep following the manufacturer's instructions. Restriction enzyme digests were used to identify likely positive clones of assembled plasmids, which were analysed via agarose gel electrophoresis. Two constructs were taken forward: pFL-GST-LIC_ADD1_RA and pFastBac-His-3C-LIC_XNP_RA for expression in insect cells. For pFastBac-XNP_RA, the plasmid is 8822bp (containing the 4332bp HIS-3C-XNP CDS), and digestion with ScaI results in 2 bands at 1519bp and 7303bp (Appendix Figure 8–9). For pFL-GST-LIC_ADD1_RA, the plasmid is 9790bp (containing the 3594bp GST-LIC_ADD1_RA CDS), and digestion with Kpn1 results in 3 bands at 342bp ,2360bp and 7088bp (Appendix Figure 8–10). Once insertion was determined, the vectors were sequenced using primers sequencing into the inserted CDS from the backbone vector (Appendix Table 20).

3.1.11 DNA sequencing

All plasmids generated using PCR amplified fragments were sequenced using Sanger sequencing with the ABI PRISM BigDye™ Terminator Cycle Sequencing kit v1.1 (Applied Biosystems) to confirm the sequence of the insert and insertion in the correct open reading frame (ORF). The sequencing reaction (12µL): 150ng Template, 2.4µL of Big Dye Terminator sequencing buffer (Applied Biosystems), 2µL of 1µM primer, 0.5µL of Big Dye Terminator enzyme mix (Applied Biosystems) and nuclease-free water (Ambion) was performed using the thermocycler program SEQFAST: 1min 96°C followed by [10s 96°C, 50s 55°C, 4min 60°C] for 40 cycles and was sent to Edinburgh Genomics for sequence analysis.

3.1.12 Glycerol Stock Preparation

Glycerol stocks were prepared by adding 500µL 50% glycerol to 500µL of an overnight culture expressing the plasmid of interest grown in 2mL of LB broth, supplemented with antibiotic (37°C shaking at 200rpm) and snap frozen in liquid nitrogen for storage at -80°C.

3.1.13 Bacmid generation and ethanol precipitation of DNA

100ng of pFL or pFastBac plasmid was transformed into 50µL EMBacY competent cells (see section 3.6.2), which use blue-white selection. DNA and cells were mixed and incubated on ice for 30min, heat shocked for 45s at 42°C and left on ice for 2min before addition of 1mL SOC. Cells were shaken at 200rpm for 4hr at 37°C then spun at 4°C for 1min at 6800 x g to concentrate the cells by removing 700µL of supernatant. Cells were then plated on EMBacY selection plates and incubated for 48 hours at 37°C. After 48hr colonies were assessed. White colonies should contain the transposed baculovirus genome, which has disrupted the lacZ gene so it cannot metabolise X-gal. Ten white colonies and two blue colonies were re-streaked on EMBacY selection plates and incubated for 48hr at 37°C, then three white colonies from the streaked plate were grown overnight in 5mL LB broth containing the same selection antibiotics as the EMBacY selection plates shaking at 37°C. The next day, cell pellets were obtained by centrifugation at 6800 x g at 4°C for 10min and the baculovirus genome recovered by miniprep following the manufacturer's instructions until the addition of N3. After neutralization, the solution was centrifuged at 18000 x

g for 10min. The supernatant was transferred to a new tube and added to 2.5 volumes of ice-cold isopropanol. DNA pellets were obtained by spinning at 10000 x g for 15min at 4°C, the supernatant removed, the pellet was then washed twice in 2 volumes of ice-cold 70% ethanol then air-dried and resuspended in 100µL of ddH_2O.

3.1.14 Synthesis of gRNAs

Guide RNA sequences were selected using the flyCRISPR Target Finder (http://tools.flycrispr.molbio.wisc.edu/targetfinder/), by insertion of the sequence of interest to be targeted. This ensured no off-targets and a prediction of cutting efficiency. sgRNAs were synthesised as gene blocks containing a dU6.2 promoter, tracrRNA and pol3 terminator. sgRNA-containing gene blocks were resuspended to a final concentration of 50µM and were amplified using dU6_2_sgRNA_F and dU6_2_3_sgRNA_R primers. Sequences of gRNAs used to target S2 cell lines are listed in Table 2.

Table 2 sgRNA gene block template names and sequences

sgRNA template name	sgRNA sequence
dU62 promoter-dPIAS_sgRNA3_IDT	AATGGTGCAGATGCTTCGAG
dU62 promoter-dPIAS_sgRNA4_IDT	CAAAAAGGGATCCAACGTAC
dU6.2_ADD1_sgRNA1_IDT	TCATGAGTAACAGTGCTCC

3.2 S2 cell culture

3.2.1 Media and growth conditions

Drosophila S2 Schneider Line 2 derived L2–4 cell line (S2) cells were cultured at 25°C at a density between $1x10^6$cells/mL- $1x10^7$cells/mL in either:
A) Schneider's *Drosophila* Medium (SERVA) (AMS Biotechnology) supplemented with 10% fetal bovine serum (FBS), 0.3mg/mL penicillin, 0.3mg/mL streptomycin and 0.75µg/mL amphotericin (B. Schneider, I. (1964) J. Exp. Zool. 156, 91-104 and 166) B) Lonza Schneider's *Drosophila* Medium Modified
C) Schneider's *Drosophila* Medium (Gibco cat. No. 21720024).

To keep cells in a growth phase, cells were passaged every 3-4 days. For S2 cell passaging, medium was brought to RT before use. For adherent cultures, the

medium was removed and the flask bashed to detach S2 cells from the plastic surface. 2mL fresh medium was added and the cells were resuspended. For routine splitting, cells were split 1:6, seeding at approximately $1x10^6$cells/mL. For experiments, cells were counted using an automated cell counter TC20 (Biorad) in Trypan Blue Dye 0.4% solution (Biorad) and plated total live cell numbers in ratios corresponding to $3x10^6$cells/ 3mL/ 25cm^2 surface area. S2 cells were transferred to a 15mL falcon tube and pellets were obtained by centrifugation (300 x g, 5min, RT in a benchtop centrifuge). The cell pellet was resuspended in 3mL fresh culture medium and seeded in a fresh flask. For experiments, after harvesting, cells were washed twice with RT PBS.

3.2.2 Cryogenic storage of S2 cell lines

Exponentially growing cells were frozen at a density of $1x10^7$cells/mL in 45% conditioned medium, 45% fresh medium and 10% dimethyl sulfoxide (DMSO)(Sigma) in aliquots of 1mL suspension per cryovial. Vials were placed in a Cryo Freezing Container (Nalgene) filled with 100% isopropanol (Fisher Scientific) and incubated overnight at -80°C, then transferred to liquid nitrogen for long-term storage.

3.2.3 Thawing S2 cell lines

Cryotubes were removed from storage and placed on dry ice for transfer to the tissue culture facility. Medium was prewarmed to 20°C and the cryovial was partially thawed in a 37°C water bath. When a small amount of ice remained, the vial was removed from the water bath, sprayed with ethanol and thawed to completion with the heat of the hand. The full contents of the tube were immediately pipetted into a 15mL falcon containing 4mL of fresh medium. The cryovial was rinsed once with 1mL medium and this added to the 15mL falcon. Cells were pelleted for 3min at 300 x g at RT and resuspended in 3mL fresh medium. For stable cell lines, antibiotics were added after the first passage.

3.2.4 Transfections in S2 cells

$5x10^5$cells/24-well plate were plated in 500µL medium the evening before transfection. For transfection reactions, 100µL serum-free medium and 200ng DNA were mixed by vortexing. 1.5µL X-tremeGENE™ HP DNA Transfection reagent

(Roche) was pipetted into the DNA-medium mix and incubated for 30min at RT then added to S2 cells in a dropwise manner. Transfections reactions were scaled in the above ratios for transfecting different surface areas of plates. 100µL of resuspended S2 cells were harvested at 3-days-post-transfection for immunofluorescence to confirm successful transfection through expression of a tag on the plasmid.

3.2.5 Generation of stable S2 cell lines

For generation of stable S2 cell lines expressing XNP-FLAG, S2 cells transfected with pMT_ATRX_FLAG_hygro were maintained in medium containing 100µg/mL Hygromycin B to retain the plasmid under selection. For induction from the metallothionein promoter (pMT), 200µM $CuSO_4$ was added to the medium three days before harvesting.

3.2.6 CRISPR/Cas9 transfection and antibiotic selection for tagged gene of interest

S2 cells were co-transfected with 190ng pIB_Cas9_CD4_Blast, 200ng homology directed repair template (HDRT) and 110ng PCR product delivering the single guide RNA (sgRNA). Details of the sgRNA and HDRT combinations to generate each cell line are listed in Table 3. 24hr post-transfection, medium was exchanged for medium containing blasticidin at 25mg/mL for selection of cells with the pIB_Cas9 plasmid, where they remained under selection for 3 days. After 3 days, medium was exchanged for fresh medium without antibiotics and cells were recovered for a further 3 days before splitting and transferring to a new tissue culture flask, taking a sample for PCR to confirm stable integration of the tag at the targeted locus after the plasmid should have been lost. Primer sequences are listed in Appendix Table 20.

Table 3 Combinations of single guide RNA (sgRNA) and homology directed repair template (HDRT) transfected and the resulting cell lines generated

	sgRNA template	HDRT	Cell line
1	dU62 promoter-dPIAS_sgRNA3_IDT	pMT_PIASNter_V5_HDRT	PIAS-N-V5
2	dU62 promoter-dPIAS_sgRNA4_IDT	pBlue_PIASC_V5_HDRT	PIAS-C-V5
3	dU6.2_ADD1_sgRNA1_IDT	pMT_ADD1_HA_HDRT	$ADD1^{HA}$

3.2.7 CRISPR/Cas9 transfection for gene destruction

S2 cells were co-transfected with 190ng pIB_Cas9_CD4_Blast and 110ng PCR product delivering the guide RNA. Knockdown was assessed by estimating the percentage of S2 cells transfected by immunofluorescsnce (IF) and, after expansion, western blotting for the protein of interest in knockdown and wild-type S2 cell lysate loaded on SDS-PAGE in equal quantities as assayed by Bradford.

3.2.8 Generation of S2 cell conditioned medium

On day one, S2 cells were plated at a density of $3x10^6$cells/mL. On day two ,medium was exchanged and on day three, S2 growing at exponential rates were harvested and the supernatant collected by centrifugation at 300 x g for 10min. This medium contains growth factors secreted by healthy S2 cells. Supernatant was passed through a 0.22μM filter to ensure removal of any cellular material.

3.2.9 S2 cell cloning

Cells were plated in conditioned medium at a low density in 96-well plates, covering a range of dilutions, aiming for single cells. After one week, colonies of more than 20 cells were picked and placed in a fresh well of conditioned medium in a 96-well plate. At this stage, clones were PCR screened for insertion of the CRISPR tag and those containing the inserted tag were expanded. Endogenously tagged cell lines were validated by PCR, IF and western blot then used for IP.

3.3 *Drosophila melanogaster* husbandry and timed embryo collection

Wild-type ORER flies were raised at 25°C on standard cornmeal medium for the duration of the study. For harvesting of embryos, newly emerging flies were collected 1-2 days post-hatching and set in cages with the number of females 5:1 times more than males. Cages were generated from grape agar plates supplemented with a dollop of yeast paste. After overnight acclimatization, grape plates with fresh yeast were changed at regular intervals in the following schedules to collect as many embryos as possible.

A) twice a day ((changed at 07:00 to collect at 11AM and 14.00) (1-3hour mixed embryos))
B) every hr from 14:00 to 19.00 (the best time for synchronised collection)

C) the plate changed at 06:00, 07:00 and 08.00 with the 07:00 plate left until 19.00 to obtain 12hr embryos.

The first plate of the day was discarded. For the duration of the synchronised embryo collection period, 6 cages were set up and cages were discarded after 4 days and fresh cages set up with newly emerged flies. All cages contained 3-7-day-old females. Plates were stored at 4°C and processed in batch. Embryos laid on grape agar plates were collected by washing grape-agar plates with embryo wash buffer and scrapping with a paintbrush into baskets sitting in embryo wash buffer. 10g of embryos were dechorionated in 50% sodium hypochlorite for 3-5min, (observing by stereomicroscope when the antlers disappeared), and thoroughly washed under flowing tap water for at least 1min to remove the sodium hypochlorite. Processing was continued for protein extraction (section 3.4.1) or fixation and immunofluorescence (section 3.4.12).

3.4 Protein protocols

3.4.1 *Drosophila* embryo extract

Dechorionated embryos were placed in 500µL embryo extraction buffer at 4°C and homogenized. The homogenate was centrifuged 3x at 20000rpm for 3min at 4°C, each time removing the upper lipid layer. The supernatant was retained and separated from yolk and cellular debris (adapted from Fromental-Ramain C, 2017[132]).

3.4.2 Protein extract from S2 cells using RIPA buffer

Cell pellets were washed twice in PBS, resuspended in 50µL RIPA buffer, incubated on ice for 30min and spun at 180000 x g for 10min. The supernatant contains the solubilized protein, 6x LB was added to give final 1x LB and boiled for 5min at 95°C.

3.4.3 Nuclear extract from S2 cells

S2 cells harvested from one T175mL flask were washed twice in cold PBS, counted and $1x10^7$cells resuspended in 400µL cell lysis buffer. The cells were then incubated on ice for 10min, passed through a 26.5G needle 12 times and incubated on ice for a further 10min. At this stage 20µL (5% total protein) for whole-cell extract (WCE) could be taken. The extract was then centrifuged at 500 x g for 5min to pellet

nuclei. The supernatant containing the cytoplasmic fraction was removed. Nuclei were resuspended in 100µL cell lysis buffer and washed once. After removal of supernatant, nuclei were resuspended in benzoase buffer and incubated on ice for 1hr, flicking the tubes occasionally. 5µL (5% total protein) was taken as a sample for nuclear extract. NaCl was added to final concentration 500mM and incubated for 30min on ice. Samples were centrifuged at 18000 x g at 4°C for 10min. Supernatant was transferred to a new tube and was quantified using a Bradford assay containing chromatin associated proteins.

3.4.4 Chromatin release assay using sequential NaCl extraction

S2 cells harvested from T175mL flasks were washed twice in cold PBS, counted and $1x10^7$cells resuspended in 400µL cell lysis buffer, washed once and resuspended in cell lysis buffer containing 0.1M NaCl and 0.1% (v/v) Triton-X100, gently flicked then incubated on ice for 5min. Cells were centrifuged at 8000 x g for 2min, and the supernatant containing the soluble sample was transferred to a new 1.5mL Eppendorf tube and the insoluble sample was resuspended in 400µL cold lysis buffer. Incubations were then repeated with buffers containing 0.1% Triton-X100 and sequentially increasing NaCl concentrations of 0.2M, 0.4M, 0.5M, 1M and 2M. After the final incubation, the cell pellet was re-suspended in 100µL cell lysis buffer with no added NaCl. 2x SDS Loading Buffer was added to all samples and the cell pellet followed by incubation at 95°C for 5min. Samples were analysed by SDS-PAGE and western blotting.

3.4.5 Protein quantification by Bradford Assay

For Bradford assays (Biorad, 5000006), dye was mixed with MQ water at a ratio of 1:4 and filtered through Whatman paper. The dilution was kept at RT. Bovine Serum Albumin (BSA) (Fisher Scientific) standards were prepared by dissolving 2mg BSA in 1mL in PBS. This stock of 2mg/mL was serially diluted, first transferring 100µL into 1mL of diluted Bradford then pipetting up and down 10 times between each transfer. Protein samples were incubated with Bradford for 5min at RT before determining the A_{595} reading on a spectrophotometer. A_{595} readings for standards were used to plot a standard curve of concentration vs A_{595}. The protein concentration in samples was determined by comparison of A_{595} to the standard curve.

3.4.6 Immunoprecipitation (IP)

After protein quantification, for each IP, 200µg of nuclear extract was diluted in high-salt buffer to a volume of 100µL and brought to 500µL with IP buffer. 25µL (5% or 7.5µg) of diluted nuclear extract was kept on ice or at 4°C for the duration of the IP to act as input.

To couple antibody to beads, 3µg antibody was added to 30µL of either Protein A Dynabeads™ (Life Technologies) for rabbit antibodies or Protein G Dynabeads™ (Life Technologies) for sheep antibodies in 0.01%Tween-PBS and rotated for 10min at RT. Beads were then washed twice with 0.01%Tween-PBS and once with IP buffer. Alternatively, for HA and V5 IPs, coupled resins were purchased, Anti-V5 agarose affinity gel (Sigma-Aldrich) and Pierce™ Anti-HA Magnetic Beads (Thermo Fisher Scientific). These were first washed in 0.01%Tween-PBS then washed and equilibrated in IP buffer before the nuclear lysate was added.

Antibody coupled beads were added to diluted nuclear lysate and incubated at 4°C on a rotator for 2hr. Beads were washed 3 times in 1mL IP wash buffer, rotated for 10min at 4°C on an overhead rotator and then washed another 3 times in IP wash buffer. After a final wash and complete removal of IP wash buffer, beads were resuspended in 30µL 1x LB and boiled at 95°C for 5min. 5µL 6x SDS-buffer was added to the input and boiled at the same time as boiling the beads for elution of immunoprecipitated material.

3.4.7 SDS-PAGE protein analysis

Protein samples were denatured in 1x SDS sample buffer for 5min at 95°C. Proteins were resolved through Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). For mass spectrometry, protein samples were denatured in 1x LDS.

6 and 8 % SDS-PAGE gels, depending on desired resolution of protein sizes, were prepared using the Bio-Rad system. Gels were run at 140V through the stacking, after which the voltage was increased to 200V and run until the dye front left the gel. Novex™ 4-20% Tris-Glycine Mini Gels were used to resolve all proteins on a full-length gel, assembled in XCell SureLock™ Mini-Cell (Thermo Fisher Scientific) and

run at 200V for 60min in 1x Novex™ Tris-Glycine SDS Running Buffer. For mass spectrometry, protein samples were run on NuPAGE™ 4-12% Bis-Tris Protein precast SDS-PAGE gels (Life Technologies) in 1x NuPAGE® MES Running Buffer (Life Technologies) through the stacking.

Either a 180 kDa PageRuler™ Prestained Protein Ladder (Thermo Fisher Scientific) or a 250 kDa Color-coded Prestained Protein Marker, Broad Range (CST) were used to estimate molecular weights.

3.4.8 Coomassie stain and Instant Blue™

After SDS-PAGE, gels were placed in dishes, washed briefly in MQ water and Coomassie stain was added to cover the gel. Gels were microwaved for 30s and rotated overnight to enhance the stain. Gels were then washed 3x10min in MQ water. For gels to be used for mass spectrometry, the gel was washed 3x10min in MQ and InstantBlue™(Expedeon) was added for 1hr, followed by 3x10min washes, before taking an image of the gel and/or extracting bands.

3.4.9 Silver stain

Silver stain was performed using the SilverQuest™ Silver Staining Kit (Invitrogen) following the manufacturer's instructions.

3.4.10 Western blot

After SDS-PAGE, proteins were transferred to Hybond® ECL™ nitrocellulose membranes (GE healthcare) by wet blotting using the Bio-Rad system by transfer in a Mini Trans-Blot Cell® at 200mA for 2hr. Following transfer, membranes were blocked in blocking buffer for 1hr at RT. The membrane was washed 3x5min in PBS-Tween20 0.1% and then incubated with primary antibody diluted in blocking buffer (see Appendix Table 18 for concentrations) overnight at 4°C. The membrane was then washed 3x5min in PBS-Tween20 0.1% and incubated with secondary antibody diluted in blocking buffer (see Appendix Table 19 for concentrations) for 45min at RT. The membrane was then washed in PBS-Tween20 0.1% and developed using 990µL luminol solution and 10µL enhancer solution for standard western blots. For increased sensitivity of detection, SuperSignal™ West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific) was used. Western blots

using HRP coupled secondary antibodies were imaged on a ChemiDoc™ Touch Imaging System (Biorad). Quantification was performed using the Quantity Tools Relative protocol in ImageLab (Biorad). The membrane was then washed in PBS-Tween20 0.1% and stored.

For LICOR antibodies, washes in PBS-Tween20 0.1% were increased to 10min after secondary incubation and the final wash performed in PBS without detergent. Membranes were visualized on the Licor Odyssey CLx instrument.

3.4.11 Ponceau

After proteins were transferred from SDS-PAGE gels to a membrane, prior to blocking, the membrane was incubated in Ponceau at RT for 5min and was washed 3x3min in 0.01%Tween-PBS to visualize protein bands.

3.4.12 Immunofluorescence and microscopy

For S2 cells, cells growing in exponential phase were resuspended and 100µL was pipetted onto poly-Lysine coated glass-slides. Poly-Lysine coated glass-slides were prepared in advance of the immunofluorescence by incubating slides (Corning) in 10% Poly-L-Lysine for 30min in a Coplin jar, then were air-dried overnight. S2 cells were left to settle for 10min in a wet chamber and were then fixed in IF fixative for 10min at RT. Slides were then washed in PBS-Triton 0.1% for 5min at RT and then blocked with Image-iT® FX signal enhancer (Thermo Fisher Scientific) for 30min at RT. Primary antibodies were diluted in goat antiserum (Life technologies) and incubated with the cells for 2hr at RT in a wet chamber or overnight at 4°C. Slides were washed 3x5min in PBS-Triton 0.1% vertically in a Coplin jar. Fluorescently labelled secondary antibodies were diluted in goat serum and incubated with the cells for 45min at RT in a wet chamber. Slides were then washed 3x5min in PBS-Triton 0.1%. The DNA was stained by adding DAPI solution and incubated for 5min at RT. Following a 2min wash in PBS-Triton 0.1%, coverslips were mounted using SlowFade® Gold Antifade Solution (Invitrogen) and sealed with nail polish. Images were acquired on a DeltaVision RT Elite Microscope and imported into ImageJ to generate overlays and then deconvoluted using softWoRx Imaging workstation (Applied Precision).

All pictures shown are quick projections of all stacks. Per image, 30-60 stacks with 0.2µm increments were taken with a 100x oil-immersion objective and a CoolSnap HQ Monochrome camera (Photometrics).

For dechorionated *Drosophila* embryos, embryos were scraped into 1.5mL Eppendorfs containing heptane:methanol in equal ratios pre-chilled at -20ºC. Eppendorfs were shaken vigorously for 60s until embryos sank to the bottom of the Eppendorf and the vitelline membrane was removed. Embryos were washed once with ice cold methanol and stored at -20°C for at least 12hr. Embryos were rehydrated in 3x10min washes in PBS-Triton 0.05% at RT. Embryos were then blocked in PBTA for 1hr at RT and then transferred to PCR tubes. Primary antibodies in PBTA were added (for concentrations, see Table 18) and rotated overnight at 4°C. The following day, the embryos were washed 3 times in PBS-Triton 0.05% and secondary antibodies (see Table 19 concentrations) in PBTA were added. Then the PCR tubes were rotated for 1hr at RT, covered in foil to maintain a light barrier. Embryos were washed 3 times in PBS-Triton 0.05% and DAPI solution was added for 5min. After a final wash in PBS-Triton 0.05%, embryos were resuspended in 30µL SlowFade® Gold Antifade Solution (Invitrogen) and were transferred to slides. A coverslip was placed over the embryos and the slide sealed with nail polish. Slides were left to air dry and then embryos were visualized using a Zeiss LSM 880 laser scanning confocal equipped with an Airyscan detector (Zeiss UK, Cambridge). Either a Plan Apochromat 20x/0.8 objective or an oil immersion Plan Apochromat 40x/1.30 objective was used for enhanced resolution.

Firstly, the top and bottom of each embryo was marked, and a capture was taken in the DAPI channel of the middle section to determine the stage of the embryo. For timed embryo collection, 100-120 embryos were staged in three independent experiments. Plotted are the stage distributions from three experiments. Error bars show the mean and standard error of mean (SEM), generated using GraphPad Prism 6 (GraphPad Software Inc.).

3.5 Mass Spectrometry Sample Preparation and Analysis

3.5.1 Mass spectrometry preparation: in-gel digest and StageTip purification

InstantBlue™ stained protein bands in SDS-PAGE gels were excised and cut into 1mm^3 sized pieces using a scalpel. Gel pieces were placed in a microcentrifuge tube, washed with 50mM ammonium bicarbonate (ABC) and 50mM acetonitrile (ACN), and shaken for 30min at 37°C. Washes with 50mM ABC/50mM ACN were carried out twice more. Proteins were reduced in 10mM DTT/50mM ABC whilst shaking for 30min at 37°C, followed by a wash with 50mM ABC and a second wash with 50mM ACN. Proteins were then alkylated in 55mM iodoacetamide in 50mM ABC for 20min at RT in the dark. After this incubation, the gel-pieces were washed in 50mM ABC/50mM ACN then incubated on ice in trypsin buffer for 15min, followed by digestion at 37°C overnight. Digestion was stopped with 1% trifluoroacetic acid (TFA) and spun through StageTips (Rappsilber *et al.,* 2003[306]). 0.1% TFA in 80% ACN was added to the gel pieces for 5min at RT and the supernatant concentrated (concentrator 5301 [Eppendorf AG, Hamburg, Germany]). 100µL of 0.1% TFA was added to the dried peptides and passed through the stage tip.

3.5.2 MALDI-TOF

The saturated matrix solution was generated from: 10mg a-cyano-4-hydroxycinnamic acid (CHCA), 400µL ddH_2O, 100µL TFA and 500µL acetonitrile, mixed and sonicated for 5min in a water bath sonicator. The matrix solution was centrifuged at 5000rpm for 1min to pellet undissolved material. Peptides were eluted from StageTips in 2µL matrix solution. 0.5µL peptide solution was mixed with 0.5µL CHCA MALDI matrix and spotted onto a stainless steel MALDI sample plate and air dried to co-crystalize.

MALDI-ToF peptide mass fingerprinting (PMF) was performed on an MALDI TOF-TOF Bruker UltrafleXtreme in the Scottish Instrumentation and Resource Centre for Advanced Mass Spectrometry (SIRCAMS) facility. Fragment mass was calibrated from a peptide calibration standard (Bruker) and mass peaks were analysed in DataAnalysis software 4.4 (Bruker). The Mascot server (http://www.matrixscience.com/server.html) was used to identify protein fragments and to assign a protein identity.

3.5.3 Mass Spectrometry (MS) LC-MS2 analysis

The digested peptides were run on a Q Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer or an Orbitrap Fusion Tribrid Mass Spectrometer by Dr. Christos Spanos (Mass Spectrometry facility). Xcalibur was used for quality control of elution spectra.

3.5.4 Peptide identification and quantification using MaxQuant

MaxQuant version 1.6.3.4 with the Andromeda algorithm was used for analysis of raw data for identification of proteins and post-translational modification sites. Peptides were searched against the *Drosophila melanogaster* reference proteome (3AUP000000803release 02.2015), downloaded from UniProt 15.11.2018.

For mapping of peptides against a reference proteome containing one entry per gene, the UP000000803 7227 DROME(https://www.ebi.ac.uk/reference_proteomes) was used. Label-free quantification (LFQ) applied. Oxidation on methionine, carbamidomethyl on cysteine and acetyl on the N-terminus of proteins were included as fixed modifications. 20ppm of mass tolerance was applied. Protein identifiers were defined by UniProt (https://www.uniprot.org) and converted into gene names.

3.5.5 Analysis of mass spectrometry data using Perseus

Perseus, (version 1.3.0.4) downloaded 13.06.2018 (https://maxquant.net/perseus/), was used to determine the protein abundance in a quantitative proteome analysis using LFQ analysis. LFQ intensity expression values were pre-processed to filter out hits with protein identifications classified as only identified by site, reverse and contaminant. The LFQ expression values were $\log_2$ transformed, categorized into groups and filtered for valid values. Protein hits identified in a minimum of two out of three biological replicates and by two or more peptides were classed as interactors.

Three types of quality control were performed on each dataset. Principal component analysis (PCA) was performed on LFQ intensity expression values from a randomly selected set of protein hits before transformation. This identified variations between the data in different conditions. The participation of the principal components in

explaining the data variance in terms of percentage is shown on the x- and y-axis of PCA plots. Multi scatter plots of log_2 transformed LFQ intensity values were generated to visualise the degree of correlation between samples, controls and replicates. The Pearson's correlation coefficient was calculated for each dataset in a pairwise comparison. A value between 0.8 and 1 shows high correlation, 0.6-0.8 weak correlation and 0.3-0.5 no correlation. Histograms were generated from LFQ intensity expression values to ensure data were normally distributed.

Missing values were imputed from normal distribution. The imputation function calculates the width and center of the distribution and shrinks the distributions to have a set standard deviation and equal variance within the groups of replicates. A two-tailed two-sample Student's T-test was used to determine whether the difference between the means of the LFQ intensity expression values between two groups (e.g. PIAS IP and IgG IP or PIAS IP -NEM and PIAS IP +NEM) was significant. The negative log p-value and the log_2 difference between the group means was plotted as a volcano plot. The threshold for significance of changes in protein abundance between the different groups (assigned p-value) was calculated using a permutation-based false discovery rate (FDR) value of 0.05 and minimal fold change value 0.1 or 2.0 (values stated in figure legends).

For comparison of protein interactors in conditions with and without NEM, the mean LFQ value of bait protein was calculated for three +NEM and three -NEM samples. LFQ values for all other interactors were then normalised to the LFQ of the bait. Scatter plots of log_2 transformed LFQ values were generated in Perseus after removing proteins identified by more than one peptide in IgG controls and filtering for proteins identified by two or more peptides in two out of three replicates.

3.5.6 GO terms analysis

A protein analysis through evolutionary relationships (PANTHER) overrepresentation test (http://pantherdb.org) against the complete GO Biological Process annotation dataset, GO-Slim Biological Process annotation dataset, GO-Slim Molecular Function annotation dataset, GO-Slim Cellular Component annotation datasets and the Reactome pathways annotation dataset for *Drosophila melanogaster* was performed using Benjamini-Hochberg FDR. Functional

classification by Protein Class was also performed. For GO term confirmation, FlyMine (flymine.org)[307] was used. FlyMine is an integrated database of genomic, expression and protein data for *Drosophila*, *Anopheles* and *C. elegans*. FlyMine also has a set of tools for conversion of protein IDs to gene IDs.

3.5.7 IntAct and Cytoscape for network analysis

Lists of Proteins identified by two or more peptides in an IP-MS were input as a query into the IntAct protein database (https://www.ebi.ac.uk/intact/). The network was exported for visualisation in Cytoscape (v.3.7.1) downloaded 03.01.2019 (http://cytoscape.org).

3.6 Protein expression and purification

3.6.1 Preparation of electro-competent cells stocks

When generating stocks of bacterial cells, streak plates on LB agar no-selection (without antibiotic), from existing glycerol stocks were made and incubated at 37°C overnight. A colony was picked and grown in 10mL LB broth shaking at 200rpm overnight at 37°C. The following day, the starter culture was added to 500mL of LB broth in a 2L flask and grown until the OD_{600} was 0.5. The culture flask was then cooled in the cold room on ice for 30min. After cooling, the culture was transferred into centrifuge tubes and the cells were pelleted at 4000 x g for 15min at 4°C in a JLA 16.25 rotor. The supernatant was discarded and the pellet of cells was resuspended in 500mL of 10% glycerol, was centrifuged, and was then re-washed in 250mL 10% glycerol, was centrifuged, and was resuspended in 20mL of 10% glycerol. After decanting the supernatant, the pellet was resuspended in 6mL of 10% glycerol. 50µL aliquots were prepared in pre-chilled Eppendorfs, snap frozen in liquid nitrogen and stored at -80°C until needed.

3.6.2 Preparation of EMBacY selection plates and competent cells stocks

EMBacY selection plates were generated by adding 1x Kanamycin, Tetracycline, Carbenacillin and Gentamycin (concentrations in the materials section, Appendix Table 17 Antibiotic stocks) to melted LB-agar (<55°C) and poured into petri dishes. Once solidified, and on the day of use, plates were supplemented with 0.5mM IPTG (10µL of 1M stock) and 40µg/mL X-gal (100µL of 20mg/mL stock) by pipetting and

spreading onto the surface of the plate. Plates were allowed to dry for 30min at 37°C.

When generating a new stock of EMBacY cells, EMBacY cells from existing glycerol stocks were streaked onto an EMBacY selection plate and incubated for 16-20hr at 37°C. A colony was picked and grown in 5mL LB broth supplemented with 1x Kanamycin, Tetracycline, Carbenacillin and Gentamycin, shaking at 200rpm overnight at 37°C. The following day, the starter culture was diluted 1:200 into 250mL of pre-warmed LB (37°C) with 20mM $MgSO_4$. Cultures were grown until the OD_{600} was 0.6, then the culture was cooled in the cold room on ice for 10min. After cooling, the culture was transferred into centrifuge tubes and spun at 4000 x g for 5min at 4°C in a JLA 16.25 rotor. Supernatant was discarded and the cell pellet resuspended by gentle pipetting to homogenize the solution in 100mL of cold TFB1. After 5min on ice, cells were pelleted by centrifugation at 1400 x g at 4°C for 10min. The supernatant was discarded and the pellet resuspended in 10mL TFB2. After 15min on ice, the cells were aliquoted into pre-chilled Eppendorfs, snap frozen in liquid nitrogen and stored at -80°C until needed.

3.6.3 Sf9 cell culture

Sf9 cells were obtained from Thermo Fisher Scientific (11496015). Upon thawing, the cells were expanded and maintained at a density between $0.5x10^6$cells/mL and $4x10^6$cells/mL in Sf-900™II SFM (Gibco) in suspension culture, shaking at 200rpm at 27°C. Growth was monitored by taking a cell count at each split and plotting growth curves over time.

3.6.4 Sf9 cell freezing and thawing

Exponentially growing cells with viability above 95% were frozen at a density of $1x10^7$cells/mL in Bambanker Cell Freezing Medium (Anachem) in aliquots of 1mL suspension per cryovial and stored at -80°C before being transferred to liquid nitrogen for long-term storage. To recover stocks from liquid nitrogen, one vial of cells was rapidly thawed in a 37°C water bath and immediately added to 20mL Sf-900™II SFM (Gibco) pre-warmed to 27°C and transferred to a shaking incubator at 200rpm for 3 days. On the third day, a 10µL aliquot was taken to assess viability by trypan blue stain. At 4 days, a second 10µL aliquot was taken and the growth

rate of the Sf9 cells determined. Sf9 cells were then either split to $0.5x10^6$cells/mL for maintenance or the Sf9 cells were pelleted and the medium exchanged.

3.6.5 Generation of viral stocks

For generation of viral stocks, cells were plated at $1.8x10^6$cells in 15mL Sf9II medium in a T75 flask and allowed to settle for 30min. Cells were transfected with 40µg of bacmid DNA in 400µL of Sf-900 II medium. 16µL X-tremeGENE™ HP DNA Transfection reagent (Roche) was pipetted into the DNA-medium mix and incubated for 30min at RT then added to S2 cells in a dropwise manner. Flasks were incubated at 27°C for 36-48hr. Yellow-fluorescent protein (YFP) expression was visualized by placing flasks under a green-fluorescent protein (GFP) filter on a fluorescence microscope to confirm successful transfection through expression of a tag on one of the bacmid plasmids. Once more than 90% of the population of cells expressed YFP (usually between 4-7 days), the supernatant (containing virus) was collected. This virus stock is designated as the V0. V0 was filtered through a 0.22µm filter and stored wrapped in foil at 4°C.

3.6.6 Small-Scale test expression in Sf9 cells

After harvesting V0, the Sf9 cells 4-7 days post-transfection were resuspended in 3 mL of Sf-900 II medium and harvested. The cell pellet was resuspended in 2mL GST purification buffer or HIS purification buffer (low imidazole), depending on the tag, and was sonicated using a Branson Digital Sonifier® (30s on/30s off at 35% amplitude for 10 cycles). The cell solution was centrifuged, and the supernatant retained as the soluble fraction. 10µL soluble (S) fraction was taken and added to 10µL of 2x SDS sample buffer. 100µL of 2x SDS sample buffer was added to the pellet for the insoluble fraction (IS). The soluble fraction was added to 30µL Pierce™ Glutathione Agarose (Thermo Fisher Scientific, cat. No. 16100), rotated at 4°C for 1hr then centrifuged to pellet the resin. The flow through was retained and a 10µL flow-through (FT) was taken and added to 10µL of 2x SDS sample buffer. The resin was washed three times in GST purification buffer or HIS purification buffer (low imidazole), depending on the tag, and 30µL 1x SDS sample buffer was added to the resin after the final wash. All samples were boiled for 5min at 95°C. SDS-PAGE followed by transfer to western blot was carried out, loading the wild-type, SOL, INSOL, PD and FT fractions to determine: firstly, if the protein was expressed;

secondly, whether it was soluble; and thirdly, if it was successfully pulled-down. Western blots were probed with antibodies against the tags (GST or HIS) and, if available, antibodies to the protein being expressed.

3.6.7 Amplification of baculovirus

Two 50mL cultures of Sf9 cells at a density of $2x10^6$cells/mL were each infected with 5mL V0 and grown at 27˚C in a shaking incubator for 3 days. After 3 days the supernatant was harvested as V1 and filtered through a 0.22µm filter and stored wrapped in foil at 4˚C. A third viral stock, V2, with increased titre was generated by adding 5mL of V1 into 100mL of Sf9 cells at $2x10^6$cells/mL, grown for 3 days and the supernatant harvested and filtered through a 0.22µm filter and stored wrapped in foil at 4˚C. After 3 rounds of viral amplification, viral supernatant had high enough infectivity for large scale infection of Sf9 cells for protein expression to produce enough protein for purification as judged by expression levels on Coomassie-stained SDS-PAGE.

3.6.8 Large-scale protein expression in Sf9 cells

Two 2L flasks each containing 800mL of Sf9 cells at a density of $2x10^6$cells/mL were infected by adding 16mL of V2 (1:50 virus:medium ratio). Sf9 cells were incubated for 2-3 days, monitoring the fluorescence and viability. Once YFP expression, expressed from the bacmid, was judged to be more than 85%, with the viability of the cells more than 90%, cells were harvested by centrifugation at 300 x g for 10min at 4°C. The cell pellet was snap frozen in liquid nitrogen and stored at -80˚C.

3.6.9 Protein purification of GST-ADD1-RA

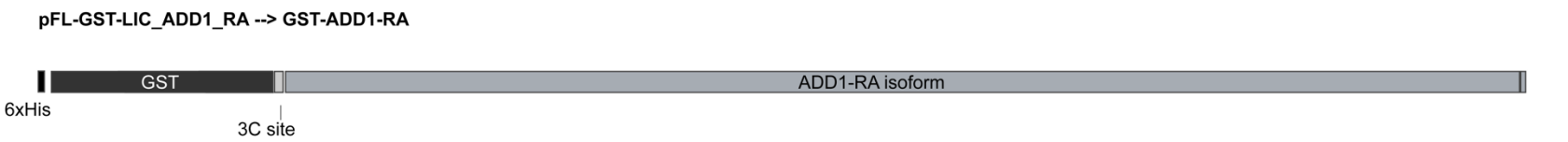
Sf9 cell pellets containing expressed GST-ADD1-RA were thawed in cold GST purification buffer (supplemented with 20µg/mL DNAse, PEFA bloc and 1x Roche cOmplete® EDTA-free Protease Inhibitors). Mg^{2+} was included in the buffer for stabilisation by providing divalent cations, DTT included to reduce oxidation damage and NaCl to maintain the ionic strength of the buffer. The cell suspension was then sonicated using a Branson Digital Sonifier® (30s on/10s off at 70% amplitude for 4 cycles). The suspension was transferred to centrifuge tubes and spun at 10000 x g for 10min at 4°C. The clarified lysate was incubated with 5mL pre-washed and

equilibrated Pierce™ Glutathione Agarose (Thermo Fisher Scientific, cat. No. 16100) for 2hr rotating at 4°C. The agarose was washed 5 times with 40mL GST purification buffer. Proteins were eluted from the agarose in 5 washes with GST purification buffer with glutathione. 250µL of 3C Protease (2mg/mL, communal stock purified from pGEX 3C made in the Cook laboratory) was added and incubated overnight, rotating at 4°C. The following day, the protein mix was concentrated using a Sartorius Vivaspin® 20 with 10,000 kDa molecular weight cut-off (MWCO) and run on size-exclusion chromatography to separate ADD1 from 3C-GST and free GST.

3.6.10 Protein purification of HIS-XNP-RA

Sf9 cell pellets containing expressed HIS-XNP-RA were thawed in cold HIS purification buffer (low imidazole) (supplemented with 20µg/mL DNAse, Pefabloc® SC (Roche) and 1x Roche cOmplete® EDTA-free Protease Inhibitors). The cell suspension was then sonicated using a Branson Digital Sonifier® (30s on/10s off at 70% amplitude for 4 cycles). The suspension was transferred to centrifuge tubes and spun at 10000 x g for 10min at 4°C. The clarified lysate was incubated with 5mL pre-washed and equilibrated HIS-Select® HF Nickel Affinity Gel slurry (Sigma, cat. No. P6611-25ML) for 2hr rotating at 4°C. An aliquot was kept as input. An aliquot of the flow- through was kept and the slurry was washed 5 times with 40mL of HIS Wash Buffer. Proteins were eluted from the HIS slurry in 5 washes with HIS purification buffer (high imidazole) and aliquots were taken and analysed by Western Blotting. The protein mix was concentrated using a Sartorius Vivaspin® 20 with 10,000 kDa molecular weight cut-off (MWCO) and run on size-exclusion chromatography. Information about the purified proteins is presented in Figure 3–1.

(A)



(B)

	ADD1		GST/3C	XNP		6HIS/3C
	whole ORF	without tag		whole ORF	without tag	
Number of bp	4332	3612	720	4017	3948	69
Number of aa	1444	1204	240	1339	1316	23
Predicted molecular weight (kDa)	**159**	**131**	28	**151**	**149**	2.9
Theoretical pI	5.49	5.29	5.18	8.40	8.58	5.43
Ext. coefficient assuming all Cys residues are reduced	97180	54320	N/A	116090	111620	N/A

Figure 3–1 Overview of recombinant protein constructs and associated information

(A) Schematic of ADD1 and XNP tagged proteins. The 3C site for protease recognition and the position of the tags are marked. (B) The table shows key parameters generated by ExPASY (http://expasy.org) after input of the amino acid sequence, which are key for designing purification strategies and for quantification of protein generated. In bold are the predicted molecular weights for tagged and untagged ADD1 and XNP.

3.6.11 Size-exclusion chromatography

All buffers, water, and 20% ethanol, were filtered through 0.22µm membrane (Millipore) and degassed prior to use on ÄKTA systems. Concentrated protein samples and nuclear extracts were centrifuged at 10000 x g for 10min at 4°C to clear particulates before injecting the protein onto a Superdex 200™10/300 GL or a Superose 6 10/300 GL column (GE Healthcare). Columns were pre-equilibrated in either HIS purification buffer (low imidazole), GST purification buffer or nuclear extraction buffer, depending on the protein being purified or experiment being performed. All equipment was provided by the Edinburgh Protein Production Facility (EPPF) and all experiments were performed in the facility. The flow rate was set at 1mL/min with 0.5mL fractions collected. Data points were exported from the UNICORN software and the $A_{280\ and}\ A_{260}$ were plotted using GraphPad Prism 6 (GraphPad Software Inc.).

3.6.12 Final protein check and generation of purified stocks

1µL 6x SB was added to 5µL taken from each fraction and boiled for SDS-PAGE analysis. The concentration of protein eluting in peaks was calculated in the UNICORN software and corroborated by measuring the A_{280} by NanoDrop 1000 Spectrophotometer (NanoDrop) by using the expected molecular weight and calculated extinction coefficient, assuming all cysteine resides were reduced. Samples were blanked against the same aged buffer as present in the eluted protein sample fraction. Based on SDS-PAGE results, fractions containing the protein of interest were combined and concentrated to 1mg/mL using pre-equilibrated Sartorius Vivaspin® 20 with 10,000 kDa molecular weight cut-off (MWCO). Aliquots of protein were snap frozen in liquid nitrogen and stored at -80°C.

3.6.13 Dynamic light scattering

Dynamic light scattering was performed on a Zetasizer Auto Plate Sampler (Malvern Panalytical) using 384 polypropylene plates (Corning) to measure particle size in suspension. 1mg/mL aliquots of protein (pre- and post-thawing) were centrifuged for 10min at 14000 rpm and diluted to 0.5mg/mL in HIS purification buffer (low imidazole) or GST purification buffer depending on the protein being purified (varying salt concentration from 0.1-0.5M NaCl). Buffer without protein was used as a blank. 50µL (25µg) was loaded per well. Each sample was measured three times following standard operating procedures as part of the Zetasizer software. All analysis was performed using Zetasizer software. The software uses a distribution analysis algorithm to give a distribution of sizes based on the decay rates contained in the correlation function.

3.6.14 Generation and testing of antiserum

The ADD1 protein was concentrated and sent to Orygen antibodies (Edinburgh) in June 2018. One rabbit was immunized, injected four times, on days 0, 25, 53 and 81, with four bleeds taken: one pre-immune serum, bleed one at day 32, bleed two at day 60, and a final bleed at day 88. Bleed two and three serum were tested by western blot and immunofluorescence. In September 2018 XNP protein was sent for immunization. Unfortunately, the company ceased operations in November 2018 and the XNP immunization schedule was terminated at day 60 after the second bleed.

Chapter 4 Identification of PIAS interactors in S2 cells

4.1 Objective

In this chapter, I present PIAS immunoprecipitation followed by mass spectrometry (IP-MS) from *Drosophila* S2 cells to identify components of the PIAS interactome. Newly available tools aided in the identification of PIAS interactors, notably, an affinity purified *Drosophila* anti-PIAS sheep antibody (a kind gift from Professor Ron Hay's lab, University of Dundee). This antibody was raised against the PIAS-RE isoform that contains only the exons included in the PIAS-CORE, which are shared by all isoforms. In addition, advances in genome editing allowed PIAS to be tagged at its endogenous locus using CRISPR/Cas9.

In addition to identifying general PIAS interactors, it was hypothesized that XNP would be a robust interactor of PIAS. With this in mind, the position of XNP and proteins associated with it are highlighted and used as a starting point to make sense of PIAS interactors in different experimental settings.

4.2 Aims

- To develop methods to isolate endogenous PIAS complexes from S2 cells to confirm or disprove that XNP is a main binding partner of PIAS
- To identify and validate other endogenous PIAS interaction partners

4.3 Results

4.3.1 Characterisation of the *Drosophila* anti-PIAS sheep antibody

The *Drosophila* anti-PIAS sheep antibody recognises epitopes in the core region of PIAS, which is shared by all PIAS isoforms. This antibody had previously been demonstrated in Ron Hay's lab to recognise two clusters of PIAS proteins on western blots and work in immunofluorescence on fixed cells. I obtained this antibody and tested it to probe western blots of lysates from cells fractionated into cytoplasmic, nuclear, chromatin associated and insoluble fractions (Figure 4–1). This confirms that PIAS is mostly in the nucleus with a portion of it chromatin associated, as previously described with a PIAS guinea pig antibody in S2 cells[141].

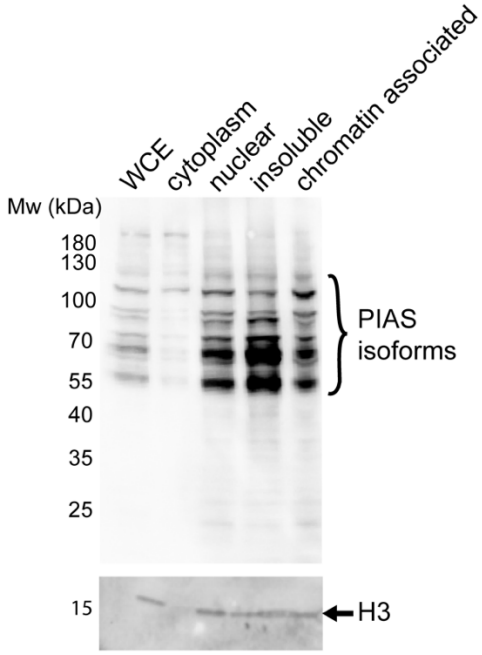


Figure 4–1 PIAS protein is concentrated in the nucleus

$1x10^7$ wild-type S2 cells were fractionated into cytoplasmic, nuclear, chromatin associated fractions and 20% of each fraction loaded. Whole cell extract (WCE), the fractions and the insoluble pellet were analysed on 4-20% SDS-PAGE (Invitrogen) followed by western blot probed with *Drosophila* anti-PIAS sheep antibody or anti-H3 rabbit antibody.

The *Drosophila* anti-PIAS sheep antibody recognises PIAS isoforms that have predicted molecular weights between 57.9 and 70.3 kDa (see Figure 2–3). PIAS-complexes can be isolated from lysate by size exclusion chromatography (Figure 4–2). These separate over a range of volumes corresponding to complexes of more than 700 kDa to less than 67 kDa (Figure 4–2), demonstrating that PIAS proteins form part of different complexes in the cell.

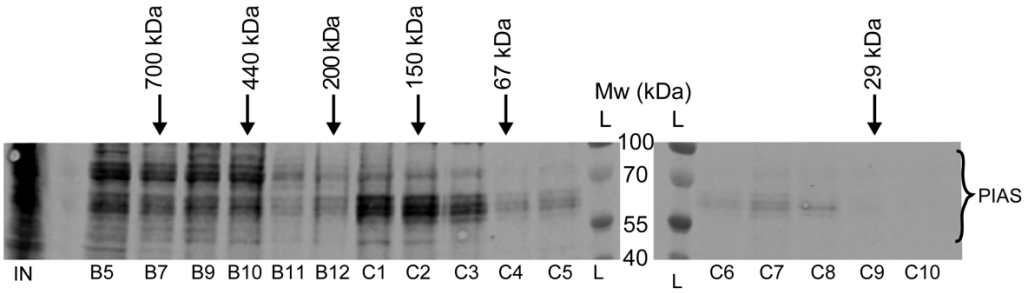


Figure 4–2 Resolution of PIAS-complexes by size exclusion chromatography
Protein extract from S2 cells separated by Superdex 200 size exclusion chromatography were analysed on 4-20% SDS-PAGE (Invitrogen) followed by western blot probed with *Drosophila* anti-PIAS sheep antibody. Elution fractions from B5-C10 are marked. The expected size of proteins in different fractions is marked.

I tested the *Drosophila* anti-PIAS sheep antibody in immunoprecipitation experiments on S2 cell nuclear extract, demonstrating that it could enrich for PIAS complexes when coupled to Protein G Dynabeads™ and immunodeplete the extract (Figure 4–3). To control for non-specific interactions, the same concentration of affinity purified IgG sheep antibody coupled to Protein G Dynabeads™ was incubated with S2 cell nuclear extract, resulting in no PIAS in the immunoprecipitated sample and presence in the flow through lane.

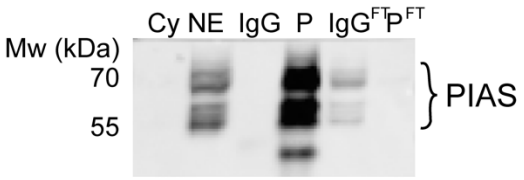


Figure 4–3 Immunoprecipitation with anti-PIAS sheep antibody immunodepletes S2 cell nuclear extract.

Protein extracts were made from wild-type S2 cells. The cytoplasmic (Cy), nuclear extract (NE), IP with affinity purified anti-IgG sheep antibody coupled to Protein G Dynabeads™ (IgG), IP with anti-PIAS sheep antibody coupled to Protein G Dynabeads™ (P) and flow through (FT) from the IgG and P IPs were analysed on 4-20% SDS-PAGE (Invitrogen) followed by western blot probed with *Drosophila* anti-PIAS sheep antibody. PIAS is enriched in the PIAS IP and is not detected in the P^{FT} lane. PIAS is not detected in IgG IP but is detected in the IgG^{FT} lane. Loading 5% input (NE), 50% IP and 5% FT.

4.3.2 PIAS interacts with HP1a

Su(var)2-10 (PIAS) has several published interactors including Su(var)205 (HP1a) (Table 4). PIAS IP followed by western blot with antibody against HP1a showed that PIAS co-immunoprecipitates this known target (Figure 4–4).

Table 4 Published physical interactors of Su(var)2-10 (PIAS)

Interactors downloaded from BIOGRID (https://thebiogrid.org). In bold is the interaction between Su(var)205 (HP1a) and Su(var)2-10 (PIAS)

Interactor A	Interactor B	Experimental System	Reference
CG7054	Su(var)2-10	Two-hybrid	Giot L (2003)
gcm	Su(var)2-10	Affinity Capture-Western	Jacques C (2009)
gcm	Su(var)2-10	Two-hybrid	Jacques C (2009)
p53	Su(var)2-10	Affinity Capture-Western	Pardi N (2011)
p53	Su(var)2-10	Two-hybrid	Pardi N (2011)
spn-A	Su(var)2-10	Two-hybrid	Formstecher E (2005)
ssx	Su(var)2-10	Two-hybrid	Giot L (2003)
stumps	Su(var)2-10	Two-hybrid	Battersby A (2003)
Su(var)2-10	CG8155	Two-hybrid	Giot L (2003)
Su(var)205	**Su(var)2-10**	**Affinity Capture-MS**	**Alekseyenko AA (2014)**

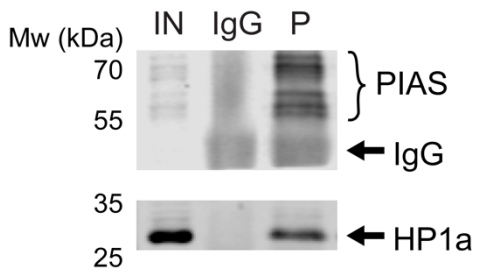


Figure 4–4 PIAS and HP1a co-immunoprecipitate.

Nuclear extracts were made from wild-type S2 cells. The input (IN), IP with affinity purified anti-IgG sheep antibody coupled to Protein G Dynabeads™ (IgG) and IP with *Drosophila* anti-PIAS sheep antibody coupled to Protein G Dynabeads™ (IP) were analysed on 8% SDS-PAGE followed by western blot probed with *Drosophila* anti-PIAS sheep antibody or anti-HP1a mouse antibody.

4.3.3 PIAS co-elutes with HP1a and ADD1 in size exclusion chromatography

To identify PIAS interactors within PIAS-complexes, *Drosophila* embryo protein lysate was separated over a size exclusion chromatography column (Figure 4–5A). Proteins in the excluded (void) volume eluted before fraction B2, while everything else is in the separation phase (after B2). Proteins that elute between 12-15mL correspond to proteins with sizes 600-700kDa (Figure 4–5A). Known heterochromatin proteins were probed for, demonstrating that HP1a, ADD1 and PIAS co-elute (Figure 4–5C) and co-immunoprecipitate from the co-eluted fractions (Figure 4–5D) with a bias for longer isoforms of PIAS co-immunoprecipitating with ADD1 (Figure 4–5C). Interaction of PIAS with other binding partners, such as XNP, was not observed by western blotting (data not shown). In fraction B2 XNP did not co-immunoprecipitate with PIAS-ADD1 complexes. This is either because the interactions are weak and have been lost during processing, or PIAS-XNP forms large complexes that are not fractionated in the separation phase.

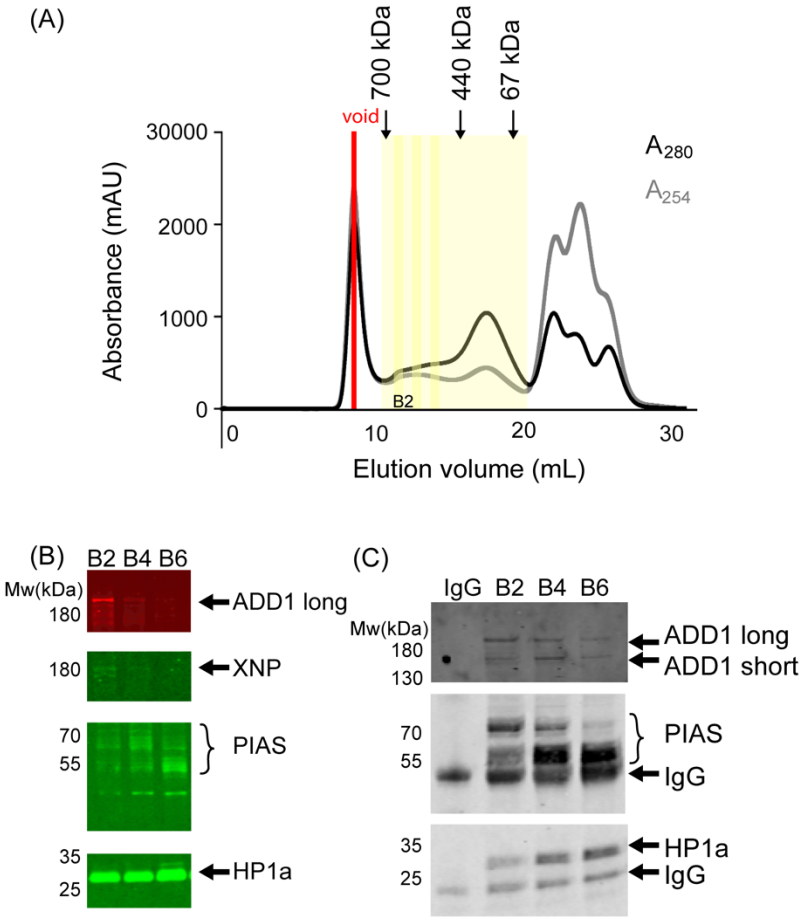
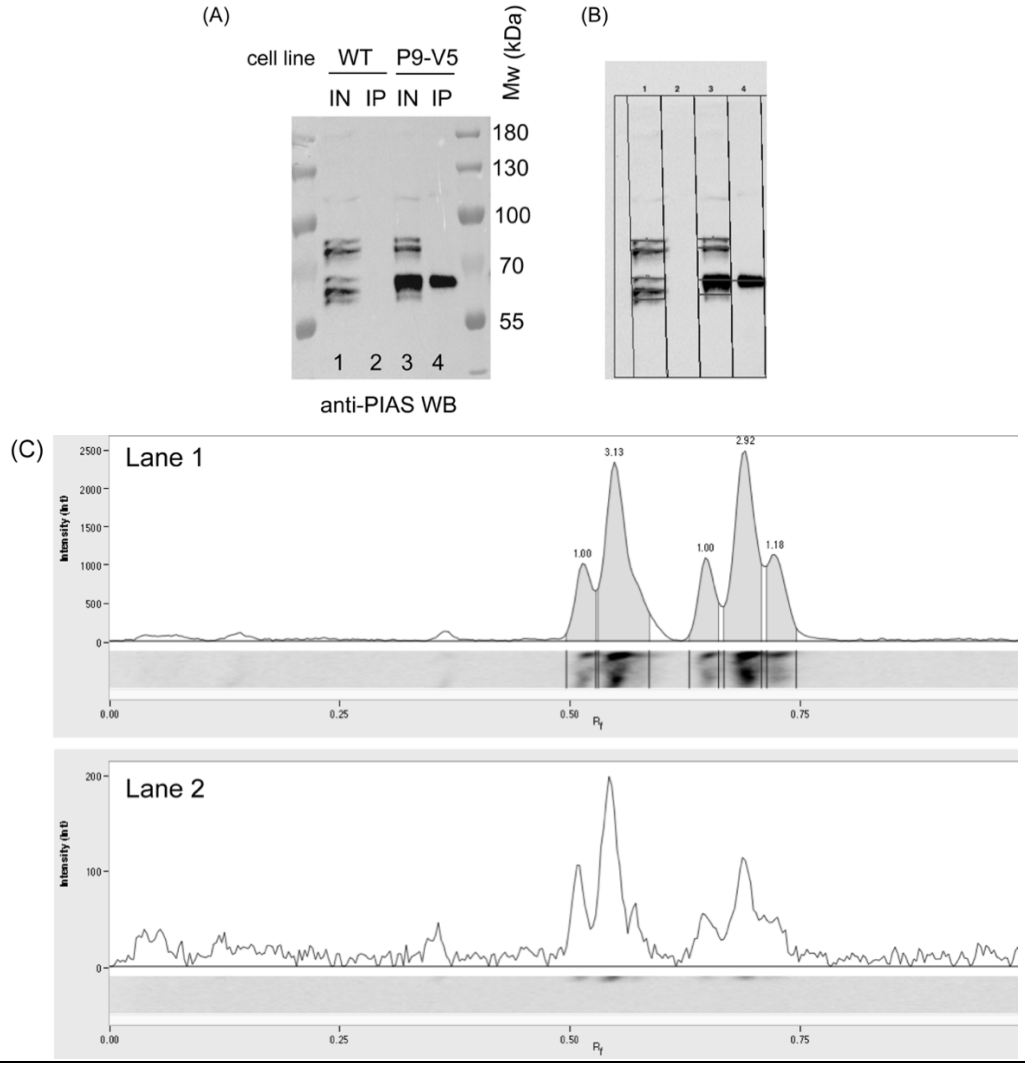


Figure 4–5 Size exclusion chromatography of *Drosophila* embryo lysate

(A) Elution profile from Superose 6 gel filtration column, in yellow is the area of protein elution collected as fractions, after the void volume of the column. The three denser yellow stripes represent the three fractions, B2, B4 and B6. The expected size of complexes that elute at a given volume are marked. **(B)** 10% of sample from fractions B2, B4 and B6 were analysed on SDS-PAGE followed by western blot probed with anti-ADD1 rabbit antibody (top panel), anti-XNP rat antibody (second panel), anti-PIAS sheep antibody (third panel) and anti-HP1a mouse antibody (bottom panel) **(C)** IP with affinity purified anti-IgG sheep antibody coupled to Protein G Dynabeads™ (IgG) and IP with *Drosophila* anti-PIAS sheep antibody coupled to Protein G Dynabeads™ from fractions B2, B4 and B6 were analysed on 4-20% SDS-PAGE followed by western blot probed with anti-ADD1 rabbit antibody (top panel) *Drosophila* anti-PIAS sheep antibody (middle panel) or anti-HP1a mouse antibody (bottom panel).

4.3.4 Characterisation of V5-PIAS-PE

Experiments performed to characterise PIAS localisation and interaction partners in S2 cells involved induced expression of V5-PIAS-PE from cells transfected with a *pMT-Pias9_V5* plasmid (the shortest isoform and closest to PIAS-CORE, Figure 2-17). This results in higher than physiological PIAS levels in the cell (Figure 4–6). To quantify levels of PIAS proteins in wild-type cells and cells expressing V5-PIAS-PE, proteins were resolved on 8% SDS-PAGE, and were analysed by western blot probing with *Drosophila* anti-PIAS sheep antibody. Quantification of PIAS levels, normalising to the largest PIAS isoform in both the top and bottom cluster of PIAS proteins, demonstrated that there was approximately 40% more PIAS protein in the extract compared to endogenous levels. For the blot in Figure 4–6, the total intensity in lane 1 is 9.23 and for lane 3 is 12.83. The difference between these is 3.65. The percentage increase is 3.65/9.23*100= 39.5%.



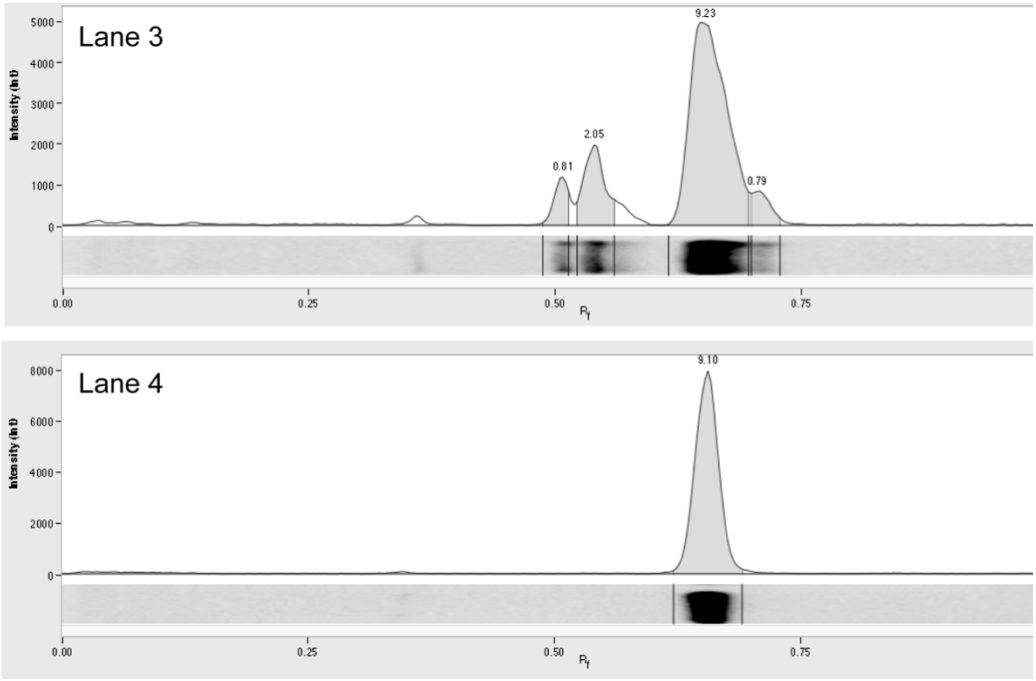


Figure 4–6 Quantitative western blot of overexpressed V5-PIAS-PE

(A) Nuclear extracts were made from wild-type Schneider S2 cells and S2 cells transfected with *pMT-Pias9_V5* in the absence of NEM. Expression of V5-PIAS-PE was induced from the *pMT-Pias9_V5* plasmid by addition of 200µM $CuSO_4$ for 3 days prior to harvesting. Cell lysate was subject to V5 immunoprecipitation using anti-V5 agarose (Invitrogen). The input (IN) and immunoprecipitation (IP) from wild-type S2 cells (WT) and from *Pias9_V5* expressing S2 cells (P9-V5) were analysed on 8% SDS-PAGE followed by western blot, and probed with *Drosophila* anti-PIAS sheep antibody followed by detected with anti-IgG sheep HRP secondary antibody. **(B)** Western blot A with lanes and shared anti-PIAS bands marked in Image Lab™ (Biorad). **(C)** Relative front (R_f) (x-axis) and Intensity (Int) (y-axis) of lanes 1-4 of the western blot in A and B. The western blot was performed by Dr. Manuella Marescotti in the Heun lab, quantified for the purpose of this thesis by myself. The experiment and the sample preparation was overseen by myself.

V5-PIAS-PE can associate and form PIAS bodies[308]. V5-PIAS-PE self-associates in puncta visible by immunofluorescence (Figure 4–7) and can be observed on western blot as a ladder of high molecular weight V5-PIAS-PE proteins (Figure 4–8). High molecular weight proteins are not present in the absence of NEM, suggesting V5-PIAS-PE forms higher MW structures promoted by SUMOylation. Previously in the Heun lab, co-transfection of pMT-Pias9_V5 and pMT_FlagHis_SUMO_GG_Hygro followed by analysis of cell lysates on SDS-PAGE and western blot demonstrated that PIAS was SUMOylated[144].

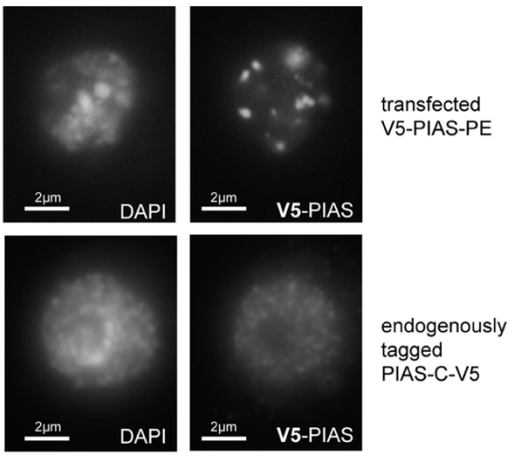


Figure 4–7 PIAS bodies form in S2 cells when PIAS is overexpressed
Immunofluorescence staining of PIAS, in a *Drosophila* S2 cell, detected by a mouse anti-V5 antibody. V5-PIAS-PE was expressed from a *pMT-Pias9_V5* plasmid or a PIAS-C-V5 CRISPR cell line was stained (see section 4.3.12).

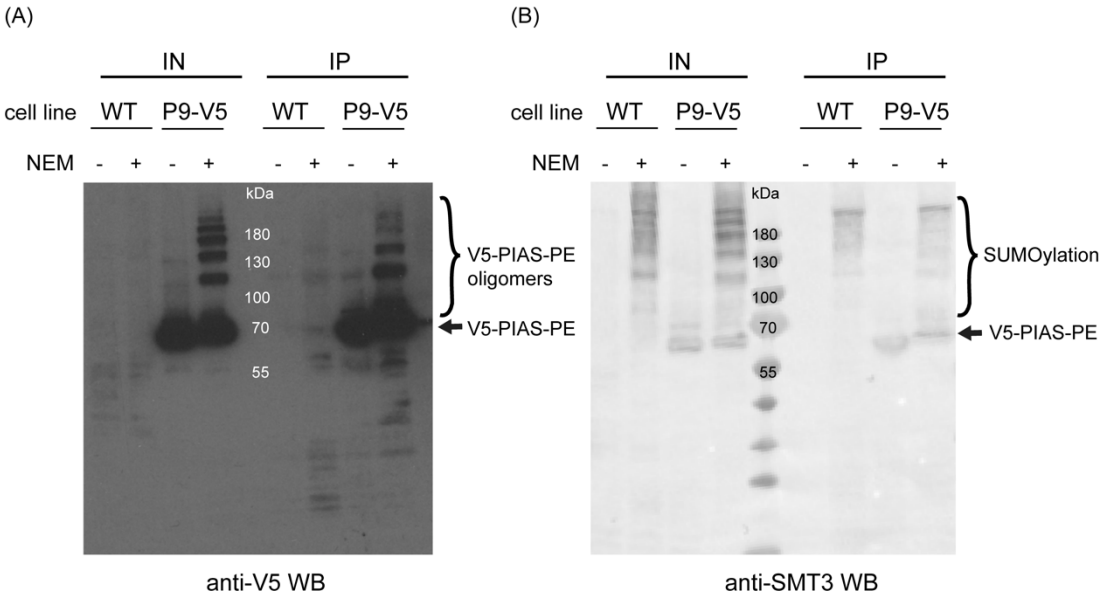


Figure 4–8 Overexpressed V5-PIAS-PE forms higher MW structures

(A) Nuclear extracts were made from wild-type Schneider S2 cells and S2 cells transfected with *pMT-Pias9_V5* in the presence (+NEM) and absence (-NEM) of NEM. Expression of V5-PIAS-PE was induced from the *pMT-Pias9_V5* plasmid by addition of 200μM $CuSO_4$ for 3 days prior to harvesting. Cell lysate was subject to V5 immunoprecipitation using anti-V5 agarose (Invitrogen). The input (IN), IP from wild-type S2 cells (WT) and IP from *Pias9_V5* expressing S2 cells (P9-V5) were anlaysed on 4-20% SDS-PAGE (Novagen) followed by western blot probed with anti-V5 mouse antibody. The bracket marks higher molecular weight species recognised by the anti-V5 antibody in +NEM conditions **(B)** Blot A re-probed with anti-SMT3 rabbit antibody. The bracket marks SUMOylated species recognised by the anti-SMT3 antibody in +NEM conditions.

4.3.5 Characterisation of endogenous PIAS

To investigate the effect of NEM on the solubility of endogenous PIAS, chromatin release assays were performed[309]. Nuclei were isolated from cells lysed in lysis buffer with and without 20mM NEM. Nuclei were treated with stepwise increases in NaCl concentration from 0.1-2.0M NaCl to sequentially extract proteins from chromatin. Proteins that form weak associations with DNA dissociate at lower NaCl concentrations than tightly bound proteins. Extracts were run on SDS-PAGE and levels of PIAS, SMT3, ADD1 and XNP assessed by western blotting (Figure 4–9).

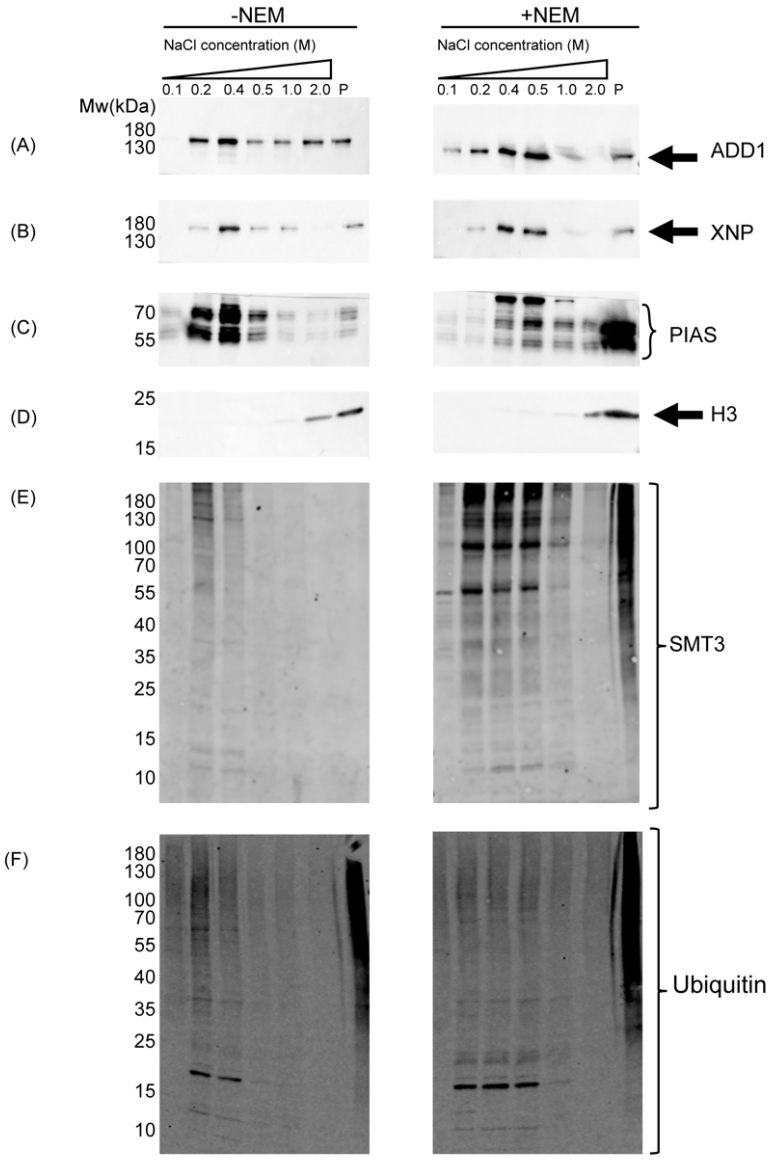


Figure 4–9 DeSUMOylation inhibitor NEM causes delayed release of PIAS in chromatin release assay.

Cells were lysed in lysis buffer with or without NEM before nuclear extraction. Nuclear extracts were washed in increasing NaCl concentrations for fractionation release from chromatin. Fractionations and solubilised pellet were anlaysed on SDS-PAGE followed by western blot probed for **(A)** ADD1, **(B)** XNP, **(C)** PIAS, **(D)** H3, **(E)** SMT3 and **(F)** ubiquitin. Release of H3 at 1.0M NaCl shows loss of chromatin structure. P= solublised pellet. Blots were cut for probing with multiple antibodies.

The presence of PIAS across the salt fractions shows sequential release of PIAS from chromatin. This suggests that different pools of PIAS are present in the cell with different solubilities. In the presence of NEM, PIAS resists extraction from chromatin even at high (2.0M) NaCl concentration, indicating that SUMOylation is responsible for chromatin association. NEM prevents the removal of ubiquitin-like substrates. SMT3 levels were higher in +NEM conditions. SMT3 was released between 0.2-0.5M NaCl (Figure 4–9), with a fraction remaining insoluble in +NEM. Although not present in the -NEM conditions, SUMOylated species are released between 0.2 and 0.5M NaCl in +NEM conditions. The extraction profile for ubiquitinated proteins was similar with and without NEM, with the majority of ubiquitinated proteins remaining insoluble in both conditions. Higher levels of ubiquitin were detected in the fraction released at 0.5M in +NEM. ADD1 and XNP have similar extraction profiles to each other and to PIAS, with more released at 0.5M NaCl in +NEM conditions (Figure 4–9). Probing for Histone H3 was used as a loading control, in addition to showing the concentration of NaCl at which chromatin is disrupted and histones are released into the soluble fraction. In this case chromatin integrity was intact up to 0.5M NaCl.

4.3.6 Affinity purification of PIAS-complexes from S2 cells

For immunoprecipitation experiments to compare PIAS interactors in the presence and absence of NEM, nuclear lysates were prepared by digestion of nucleic acids with 150U benzoase for 1hr on ice in low salt conditions (20mM). This was followed by salt extraction in 500mM NaCl, which solubilises PIAS in both the presence and absence of NEM (Figure 4–10).

Like V5-PIAS-PE (Figure 4–8), endogenous PIAS can associate and form higher molecular weight species which are maintained only in the presence of NEM (Figure 4–10). In wild-type S2 cells, endogenous PIAS forms stable complexes that resist denaturation by boiling in SDS and are visible on western blot probing with the *Drosophila* anti-PIAS sheep antibody. This implies that these are very tight non-covalent interactions whereas it is more likely that the higher Mw species are covalently attached to SMT3. These bands are visible in the presence of NEM but not visible in its absence (Figure 4–10). In the presence of NEM, there is a difference in the PIAS western blot profile with fewer clusters of lower molecular weight proteins (Figure 4–10).

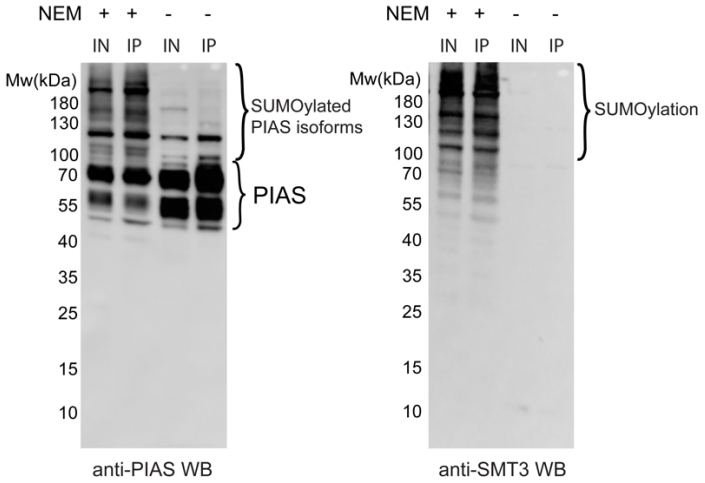


Figure 4–10 NEM treatment results in oligomerisation of PIAS and these oligomers can be immunoprecipitated using the *Drosophila* anti-PIAS sheep antibody.
Nuclear extracts were made from wild-type S2 cells in the presence (+NEM) and absence (-NEM) of NEM. The input (IN) and material immunoprecipitated with *Drosophila* anti-PIAS sheep antibody coupled to Protein G Dynabeads™ (IP) were analysed on 4-20% SDS-PAGE (Invitrogen) followed by western blot probed with *Drosophila* anti-PIAS sheep antibody or anti-SMT3 rabbit antibody.

4.3.7 Mass spectrometry of affinity purified PIAS-complexes from S2 cells

After optimising the conditions, immunoprecipitation followed by mass spectrometry (IP-MS) of PIAS, was performed to identify proteins that co-immunoprecipitate with PIAS in the absence and presence of NEM. The elution from the control IgG and PIAS IPs were visualised on silver stained SDS-PAGE gels (Appendix Figure 8–3) showing proteins specific to the PIAS IP are absent in the IgG control.

Over 1400 proteins were identified in a trial of PIAS IP-MS (Figure 4–11), with more peptides and proteins identified in NEM conditions. A large proportion of these hits were cytoskeletal, involved in translation initiation and involved in transport of material from the nucleus to the cytoplasm. Datasets were compared to see which proteins immunoprecipitated with PIAS in the presence and absence of NEM. This allowed identification of PIAS interactions that are dependent upon SUMOylation. 1400 proteins were detected in -NEM and 1800 in +NEM. The reference FASTA contained 41,655 protein entries meaning that PIAS interacts with approximately 4-5% of the *Drosophila* proteome. As many proteins were identified, further optimisation of IP conditions was performed.

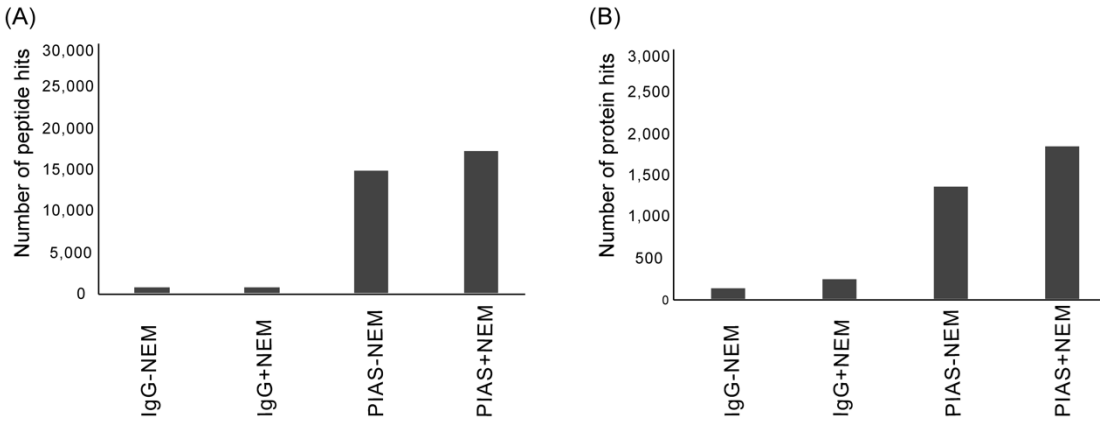


Figure 4–11 Number of peptide and protein hits from immunoprecipitation of PIAS-associated complexes

Nuclear extracts were made from wild-type S2 cells in the presence (+NEM) and absence (-NEM) of NEM. The material that co-immunoprecipitated with *Drosophila* anti-PIAS sheep antibody or anti-IgG sheep antibody coupled to Protein G Dynabeads™ (IP) was analysed by mass spectrometry (IP-MS). Numbers of peptide **(A)** and protein **(B)** hits are plotted. This is the result of one experiment.

Two methods were used to increase the stringency of the IP: firstly, increasing the number of washes from five to eight and secondly, increasing the NaCl concentration from 0.1M to 0.5M in the IP wash buffer.

After five washes, the majority of unbound protein was removed. Increasing to eight washes, as visualised by silver stain, did not result in further protein loss (data not shown). Initially IP washes used low salt concentrations (100mM) to preserve interactions, but washes in IP wash buffer containing 300mM and 500mM NaCl were also tested. In the presence of NEM the number of peptides and proteins identified remained similar with higher salt washes (Figure 4–12). This indicates robust interactions between PIAS and interactors when NEM is present, which could be due to maintenance of post-translational modifications or the nature of SUMO-SIM interactions remaining unaffected by increasing salt concentrations. However, in the absence of NEM, fewer peptides and proteins were identified when the PIAS-IPs were washed in IP wash buffer containing higher NaCl concentrations (Figure 4–12). IP washes in IP wash buffer containing 500mM NaCl emphasised the difference in interaction strength between PIAS and its interaction partners in the presence and absence of NEM (Figure 4–12). Also, using 500mM NaCl in washes, the bait bound to antibody-bead complexes was not compromised and increased the amount of PIAS detected, increasing the stringency of the IP-MS.

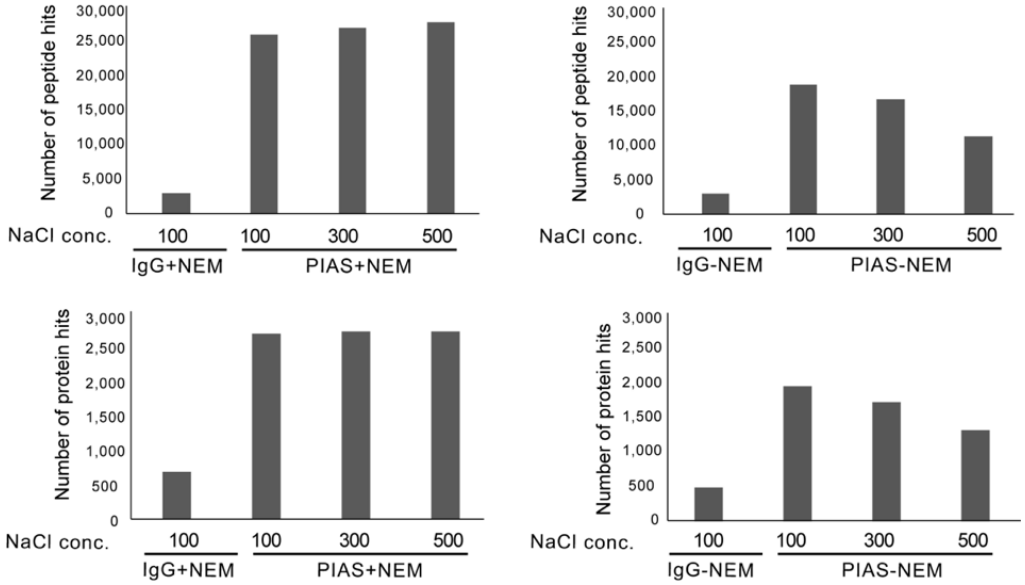


Figure 4–12 Number of peptide and proteins hits from immunoprecipitation of PIAS-associated complexes after varying the concentration of salt in IP washes

Nuclear extracts were made from wild-type S2 cells in the presence (+NEM) (left) and absence (-NEM) (right) of NEM. The material that co-immunoprecipitated with *Drosophila* anti-PIAS sheep antibody or anti-IgG sheep antibody coupled to Protein G Dynabeads™ was washed with IP wash buffer containing either 100, 300 or 500mM NaCl. The IP samples were then analysed by mass spectrometry. Numbers of peptide (top) and protein (bottom) hits are plotted. This is the result of one experiment.

Following PIAS IP and western blot, XNP was probed for as an example of a protein known to be SUMOylated and to interact with PIAS[171]. Western blotting for XNP in PIAS immunoprecipitations showed XNP is retained in PIAS IPs when washed with IP wash buffer containing 0.5M NaCl in the presence of NEM, but is not retained in the absence of NEM (Figure 4–13), suggesting that the XNP-PIAS interaction is SUMO dependent.

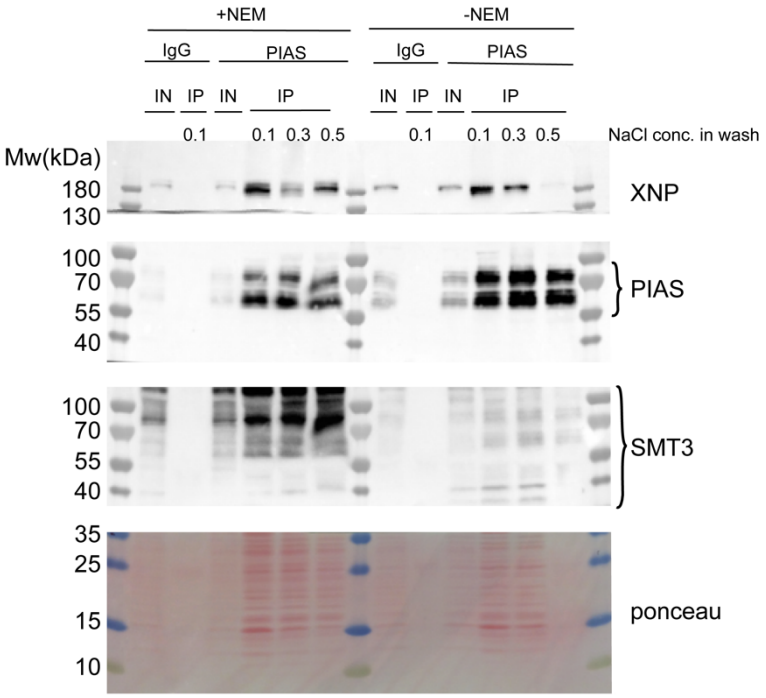


Figure 4–13 PIAS co-immunoprecipitates XNP in an NEM-dependent manner
Nuclear extracts were made from wild-type S2 cells in buffers with or without NEM. IP with affinity purified anti-IgG sheep antibody coupled to Protein G Dynabeads™ (IgG) and IP with anti-PIAS sheep antibody coupled to Protein G Dynabeads™ (IP) were washed with IP wash buffer containing 0.1, 0.3 or 0.5M NaCl. Input (IN) and immunoprecipitation (IP) were analysed on 4-20% SDS-PAGE followed by western blot. The blot was cut and blotted for XNP, PIAS and SMT3. The Ponceau for the bottom half of the blot is shown, showing reduction in protein in 0.5M NaCl IP wash in -NEM PIAS IP lane without loss of efficiency of PIAS immunoprecipitation.

PIAS and IgG IPs were performed on protein extracted from $1x10^7$ S2 cells from three biological replicates, processed in parallel, with eight washes of 0.5M NaCl prior to elution from beads. Samples were prepared for mass spectrometry and were run with identical instrument settings. Label free quantitation (LFQ) was used as a measure of the abundance of proteins. PIAS (the bait) had the highest LFQ values validating that the IP was efficient and specific (Figure 4–15).

4.3.8 PIAS isoform identification by mass spectrometry

Peptides mapping to unique amino acid sequences in PIAS isoforms: L/J, I/G/B, H/E and A/C/D were identified in different PIAS IP-MS runs, confirming immunoprecipitation of multiple PIAS isoforms with the *Drosophila* anti-PIAS sheep antibody. Most peptides map to the shared core region of PIAS, which makes up the majority of the PIAS sequence (Figure 4–14). In S2 cells most peptides mapped to PIAS isoform L (Figure 4–15). Some peptides mapped to PIAS isoform H, as these included a sequence at the end of the core region not present in isoform L (Figure 4–14, Figure 4–15). It would not be possible to comment on enrichment of particular isoforms, only their presence, due to few peptides mapping to unique isoform specific sequences. Only PIAS isoforms L and H were included in data analysis from PIAS- IP-MS in S2 cells as only these isoforms were identified by two or more peptides in two out of three replicates.

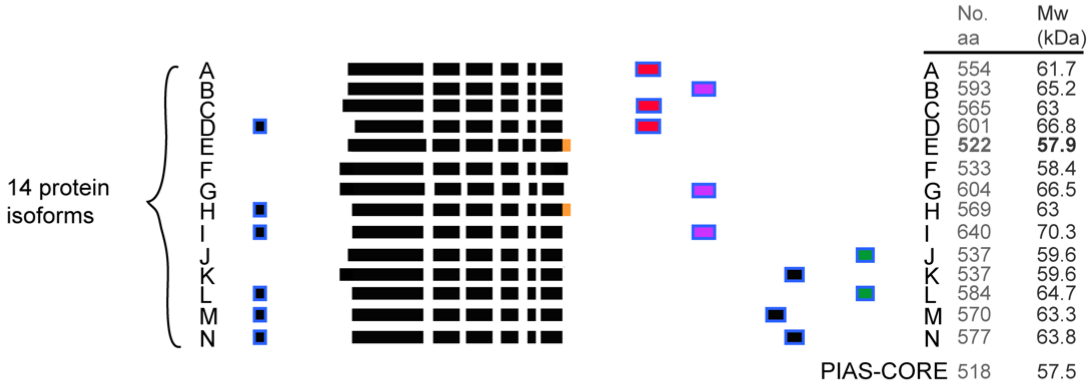


Figure 4–14 PIAS isoforms highlighting C-terminal exons shared between isoforms. 14 PIAS isoforms (A-N), exons in the core region are shown as black boxes, coloured boxes represent sequences not present in all isoforms. To the right are the number of amino acids (No.aa) and molecular weights (Mw) of each PIAS isoform.

4.3.9 Mass spectrometry analysis of PIAS interactors in S2 cells

Scatter plots of LFQ values of each protein identified in IP-MS samples were plotted against all other samples and controls. Pearson correlation values over 0.9 between replicates of IP with anti-PIAS sheep antibody coupled to Protein G Dynabeads™ in +NEM and -NEM conditions showed reproducibly between samples (Appendix Figure 8–4). Correlation between IP with affinity purified anti-IgG sheep antibody coupled to Protein G Dynabeads™ negative controls indicates that common contaminants are present, but there is no correlation between samples and controls. A principle component analysis (Appendix Figure 8–5) showed ~90% of variation in the dataset could be summarised by the difference between IP with anti-PIAS sheep antibody coupled to Protein G Dynabeads™ and IP with affinity purified anti-IgG sheep antibody coupled to Protein G Dynabeads™, indicating that the largest difference is between groups and not within the data itself.

Volcano plots were generated using Perseus to visualise the result of a Student's T-test comparing interactors identified in PIAS-IP and IgG-IP conditions after filtering for proteins identified in at least two out of three replicates and identified by two or more peptides per protein (Figure 4–15). Proteins identified by two or more peptides in the IgG control were excluded from analysis. Also plotted are proteins identified by 20 or more peptides (Figure 4–16) as this increases confidence that these are genuine PIAS binding partners.

This page is intentionally left blank

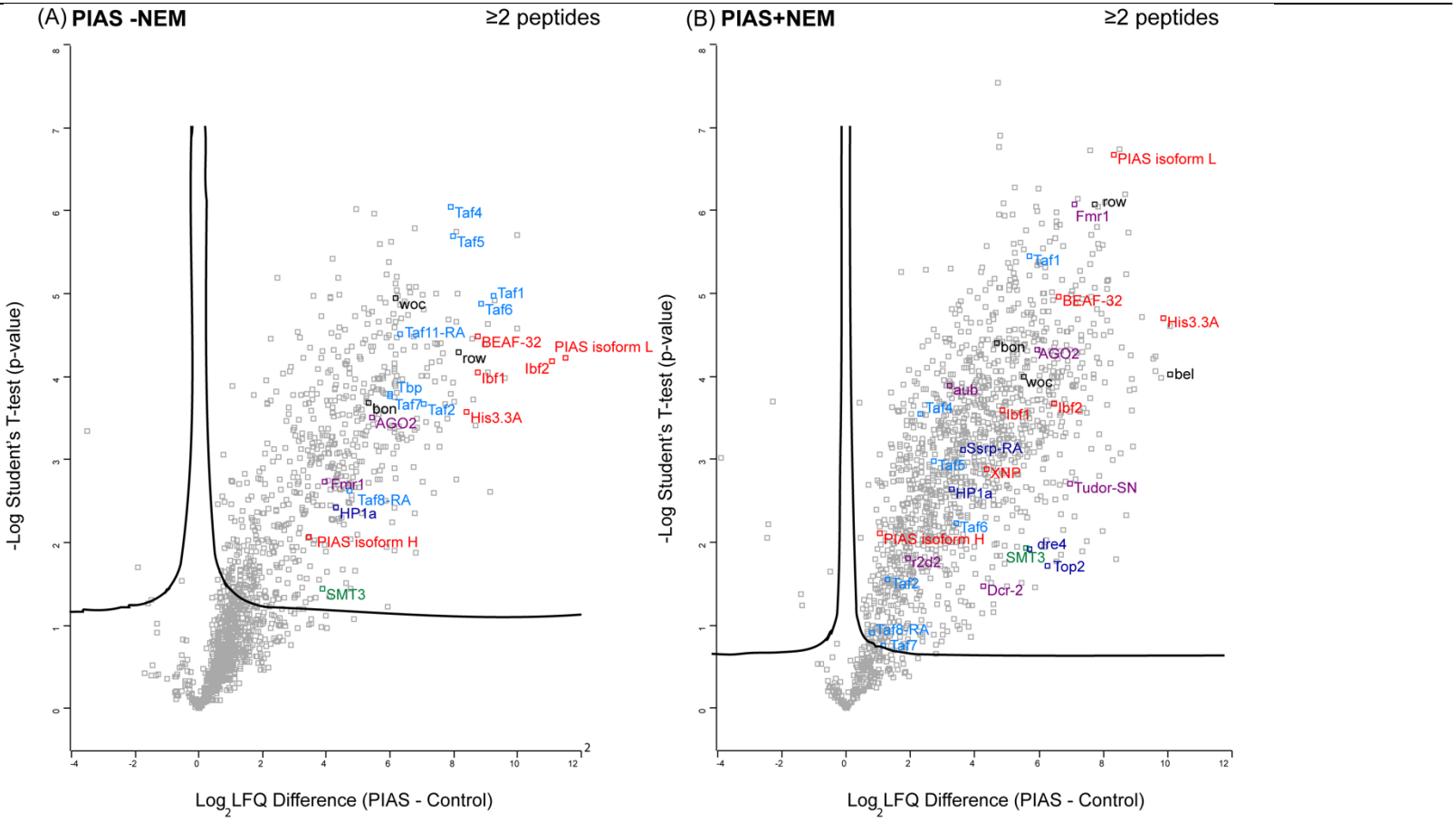

(A) PIAS -NEM
≥2 peptides
-Log Student's T-test (p-value)
Log2 LFQ Difference (PIAS - Control)
Taf4
Taf5
Taf1
Taf6
woc
Taf11-RA
BEAF-32
PIAS isoform L
row
Ibf2
Ibf1
Tbp
Taf7
Taf2
bon
His3.3A
AGO2
Fmr1
Taf8-RA
HP1a
PIAS isoform H
SMT3
(B) PIAS+NEM
≥2 peptides
-Log Student's T-test (p-value)
Log2 LFQ Difference (PIAS - Control)
PIAS isoform L
row
Fmr1
Taf1
BEAF-32
His3.3A
bon
AGO2
bel
woc
aub
Taf4
Ibf1
Ibf2
Ssrp-RA
Taf5
XNP
Tudor-SN
HP1a
Taf6
PIAS isoform H
dre4
SMT3
Top2
2d2
Taf2
Dcr-2
Taf8-RA
Taf7

Figure 4–15 Volcano plot of PIAS protein interactors in S2 cells in the absence and presence of NEM, two or more peptides

Immunoprecipitation (IP) using *Drosophila* affinity purified anti-PIAS antibody (PIAS) or control sheep anti-IgG antibody (Control) coupled to Protein G Dynabeads™ was performed from S2 cell lines. IP samples were prepared as described in methods and analysed by mass spectrometry. PIAS interactors in conditions without NEM (**A)** and with NEM **(B)** are plotted. Each dot represents one protein. For determining the PIAS interactome, a MaxQuant label-free quantitation pipeline was used. Student's T-test was performed on log_2 transformed LFQ values from three control and three PIAS IPs after filtering for proteins identified by two or more peptides and found in at least two of three replicates. T-test difference ratios (x-axis) were plotted against the negative logarithmic p-value of the T-test (y-axis). Proteins in red are proteins of interest to this PhD study including PIAS, BEAF-32, Ibf1, Ibf2, ADD1, XNP and H3.3. Proteins in purple are linked to RNAi silencing pathways. Proteins in black are consistent PIAS interactors. Proteins in blue are transcription activation proteins (Tafs). SMT3 is in green and HP1a, dre4 and Ssrp-RA are in navy blue. The confidence line was generated using a permutation-based FDR value of 0.05 and a minimal fold change value of 0.1.

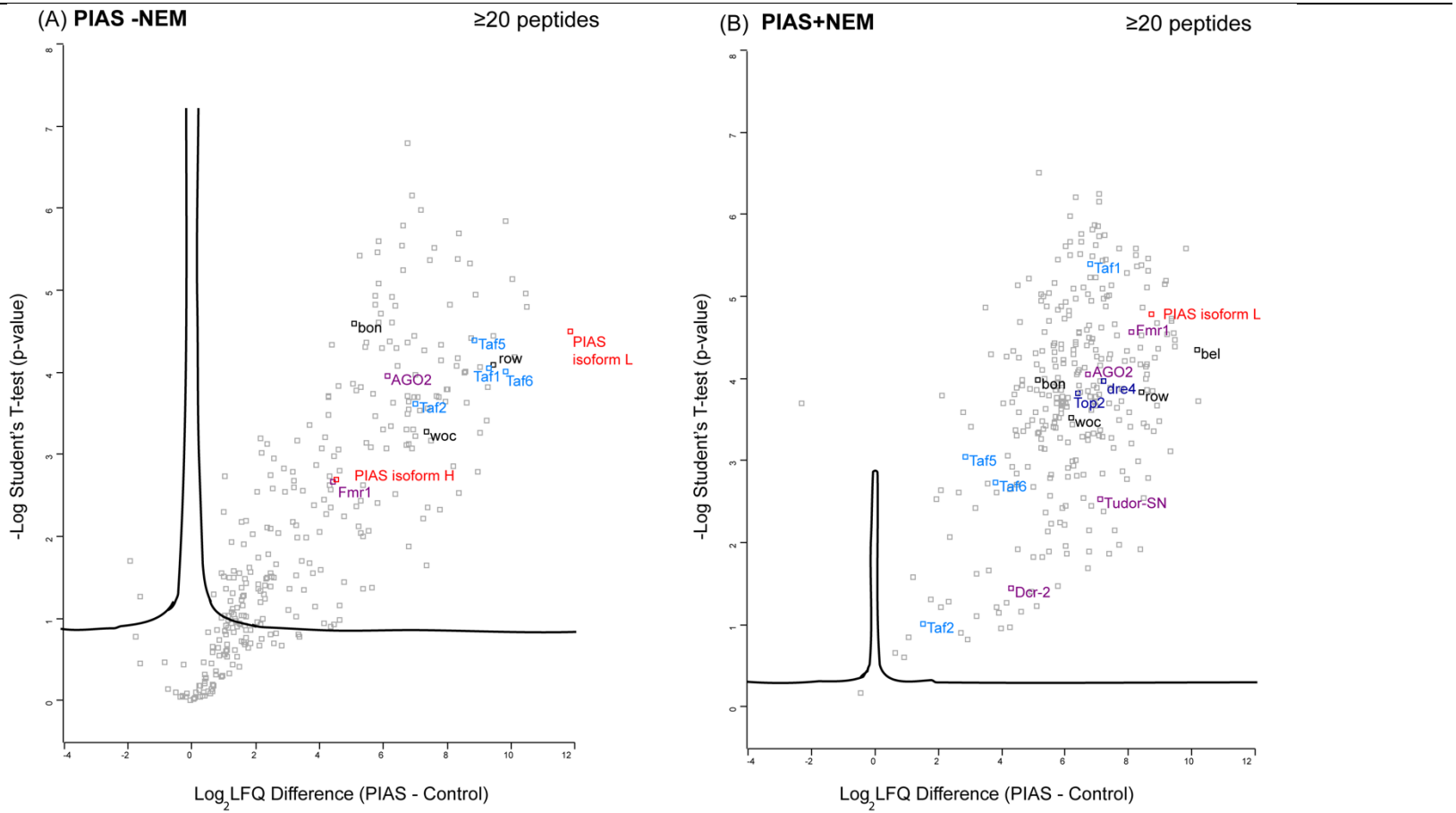

(A) PIAS -NEM
≥20 peptides
-Log Student's T-test (p-value)
Log2 LFQ Difference (PIAS - Control)
boh
Taf5
row
PIAS isoform L
AGO2
Taf1
Taf6
Taf2
woc
PIAS isoform H
Fmr1
(B) PIAS+NEM
≥20 peptides
-Log Student's T-test (p-value)
Log2 LFQ Difference (PIAS - Control)
Taf1
PIAS isoform L
Fmr1
bel
AGO2
boh
dre4
row
Top2
woc
Taf5
Taf6
Tudor-SN
Dcr-2
Taf2

Figure 4–16 Volcano plot of PIAS protein interactors in S2 cells in the absence and presence of NEM, twenty or more peptides
Immunoprecipitation (IP) using *Drosophila* affinity purified anti-PIAS antibody (PIAS) or control sheep anti-IgG antibody (control) coupled to Protein G Dynabeads™ was performed from S2 cell lines. IP samples were prepared as described in methods and analysed by mass spectrometry. PIAS interactors in conditions without NEM (**A)** and with NEM **(B)** are plotted. Each dot represents one protein. For determining the PIAS interactome, a MaxQuant label-free quantitation pipeline was used. Student's T-test was performed on log_2 transformed LFQ values from three control and three PIAS IPs after filtering for proteins identified by 20 or more peptides and found in at least two of three replicates. T-test difference ratios (x-axis) were plotted against the negative logarithmic p-value of the T-test (y-axis). PIAS is in red. Proteins in purple are linked to RNAi silencing pathways. Proteins in black are consistent PIAS interactors. Proteins in blue are transcription activation proteins (Tafs). Dre4 is in navy blue. The confidence line was generated using a permutation-based FDR value of 0.05 and a minimal fold change value of 0.1.

In this one step purification, a large number of interactors were discovered, with many involved in RNA polymerase II transcription, RNA processing and chromatin binding. PIAS also interacts with components of the RNAi machinery (AGO2, Ars2, Fmr1, r2d2, Rm62, Tudor-SN, bel, Dcr-2, aub) and Tafs. This is compatible with PIAS being involved in the transcription of repetitive elements for silencing of heterochromatic regions as previously proposed in the Heun lab[144].

Focussing on proteins identified by two or more peptides, one of the most enriched proteins, by more than eight-fold in the PIAS IP compared to control, is histone 3.3 (H3.3) (Figure 4–15). H3.3 is incorporated into nucleosomes at telomeres, pericentromeres and transposable elements by its chaperone DAXX, aided by the ATRX chromatin remodeller. In *Drosophila* DAXX-like protein (DLP) interacts with XNP and has two verified SIMs[132]. DLP was not identified as a PIAS interactor in this IP-MS, but interaction between PIAS and H3.3 could mediate SUMO dependent incorporation into chromatin. Notably, XNP was identified in +NEM conditions, enriched four-fold over background (Figure 4–15).

Of the significant proteins identified by 20 or more peptides, 230 were identified in -NEM and 313 in +NEM IP conditions (Figure 4–16). PIAS interacts with Tafs, consistent with a role for PIAS in regulation of transcription and in agreement with interactions found for human PIAS proteins in two-hybrid assays[310]. The transcription factors, without children (woc) and relative of woc (row) were identified. Both interact with the HP1c complex at developmentally regulated and stimulus-responsive genes to regulate RNA polymerase II pausing at transcription start sites[311].

Woc and row co-purify with Boundary Element Associated Factor 32 (BEAF-32) which acts as an insulator, binding near transcription start sites (often of housekeeping genes)[312]. Knockdown of BEAF-32 has minimal effects on topologically associated domain (TAD) organisation[313] characteristic of redundancy among insulator proteins. Insulator proteins (Ibf1 and Ibf2) were also identified as PIAS interactors. PIAS and BEAF-32 bind at similar sites in the genome[57] consistent with a role for PIAS mediated SUMOylation in co-ordinating transcription units.

Other proteins that co-immunoprecipitated with PIAS, but are not highlighted on volcano plots, include proteins involved in mRNA splicing and the spliceosome. Of the 313 proteins significantly enriched and identified by more than 20 peptides, 218 had annotated GO terms in FlyMine. 132 of these proteins were involved in nucleic acid metabolism: 51 in RNA metabolism and 40 in splicing. The transcription elongation factor spt5 was more than 10-fold enriched in PIAS IP over control IgG IP. Spt5 is involved in co-transcriptional spliceosome assembly with the pre-mRNA processing protein Prp8 at intron containing genes[314]. Prp8 was also over eight-fold enriched in PIAS IP over control IgG IP. For the role of SUMOylation in splicing see section 2.3.10.

Of the published *Drosophila* PIAS interactors (Table 4), only HP1a was present in this IP-MS. The position of the protein Bonus (bon) is marked as this is responsible for HP1a phosphorylation. Many proteins identified in the PIAS IP-MS are known to be SUMOylated. For example, CENP-C which was identified by 38 peptides in +NEM conditions.

Normalisation of the LFQ values of proteins that co-immunoprecipitate with PIAS allowed a comparison of PIAS interactors in the absence and presence of NEM (Figure 4–17). A correlation plot of PIAS interactors comparing the mean $\log_2$ transformed LFQ values from three biological replicates in PIAS IP-MS in + and - NEM conditions was generated to identify PIAS interactors unique to NEM conditions. 732 proteins interacted with PIAS only in the presence of NEM, 710 proteins interacted with PIAS in both the presence and absence of NEM and 135 proteins were identified only in conditions when NEM was not present (Figure 4–17). Of the 710 proteins enriched in both PIAS IP-MS conditions, in the presence and absence of NEM, 92 proteins differed less than two-fold. A Gene Ontology (GO) terms analysis of these interactors found chromosome organisation to be the most significant biological process that these factors were involved (Table 5). 34 factors including ADD1, Cenp-C and BEAF-32 were found in this category. In Table 6 are the proteins which do not significantly change between PIAS IP in conditions with and without NEM categorised into groups based on their biological processes. Interestingly, analysis of the 92 proteins by FlyMine identified an enrichment for

proteins with a zinc-finger, PHD-type domain. These include: enok, Cfp1, CG1815, Br140, bip2, rno, UPset and ADD1. FlyMine also identified pathway enrichment for RNA polymerase II transcription, transport of mature mRNA derived from an intronless transcript and regulation of TP53 Activity.

Of the proteins identified to interact with PIAS in both + and -NEM conditions, 614 proteins were identified by LFQ values greater than two-fold enriched in +NEM conditions compared to three proteins identified by LFQ values greater than two-fold enriched in -NEM conditions (Figure 4–17), possibly suggesting SUMO presence is involved in stabilising protein complexes. 732 vs 135 proteins interact with PIAS in the presence of NEM compared to when NEM is not present (Figure 4–17). In PIAS +NEM conditions, SMT3 is enriched more than two-fold and components of the RNAi machinery (Rm62, bel, Tudor-SN, Dcr2, Fmr1) are also enriched more than two-fold (Figure 4-17).

GO terms analysed against the GO-Slim Biological Process annotation dataset, GO-Slim Molecular Function annotation dataset, GO-Slim Cellular Component annotation datasets and the Reactome pathways annotation dataset was performed for the 732 proteins that interacted with PIAS only in the presence of NEM (Table 7) and the 135 proteins that were only identified in the absence of NEM (Table 8). The GO-Slim terms highlight that the common properties of proteins identified to interact with PIAS in the absence of NEM are involved in transcription and gene expression (Table 8). In contrast in the presence of NEM, where SUMOylation is maintained, a greater number of processes were identified, with a common theme of involvement of PIAS associated complexes in general nucleic acid metabolism (Table 7).

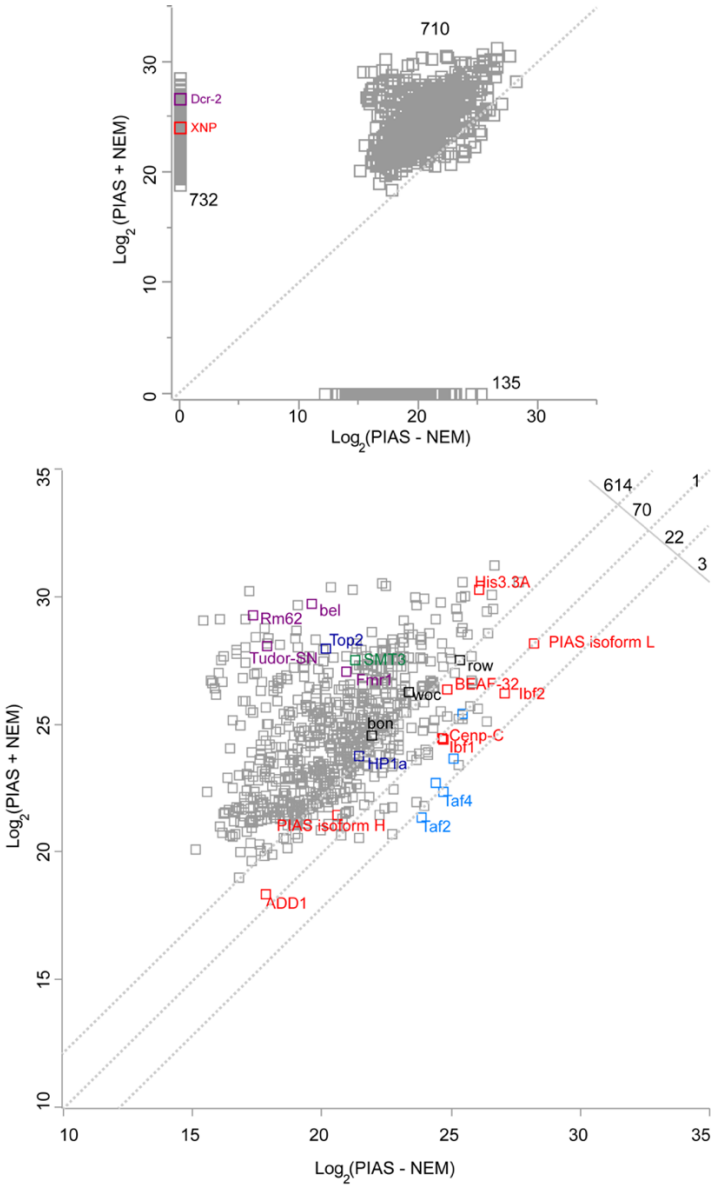


Figure 4–17 Scatter plot comparing PIAS protein interactors in S2 cells in the absence and presence of NEM

Mean log_2 transformed LFQ values calculated from three biological replicates, normalised to PIAS isoform L, from PIAS IP in +NEM and -NEM conditions. Each dot represents one protein. The top plot shows all proteins and the bottom plot shows proteins found only in both conditions. Proteins in red are of interest to this PhD study including PIAS, BEAF-32, Ibf1, Ibf2, ADD1, XNP and H3.3. HP1a is in navy blue. Proteins in purple are linked to RNAi silencing pathways. Proteins in black are consistent PIAS interactors. Proteins in blue are Tafs. SUMO (SMT3) is in green. Diagonal lines mark two log_2 fold change from zero.

Table 5 Gene ontology (GO) terms associated with proteins identified in PIAS IP-MS from S2 cells in conditions with and without NEM

Gene Ontology analysis using the FlyMine complete GO biological process annotation dataset of proteins identified in PIAS IP-MS from S2 cells with less than two log_2 fold change with and without NEM. Bonferroni test correction was applied. Matches are unique proteins within a given GO term. List is ordered by significance p-value.

GO term: Biological Process	p-value	matches
chromosome organization [GO:0051276]	8.82E-17	34
peptidyl-lysine modification [GO:0018205]	7.09E-10	15
chromatin organization [GO:0006325]	1.52E-09	23
histone modification [GO:0016570]	2.54E-09	16
covalent chromatin modification [GO:0016569]	3.22E-09	16
organelle organization [GO:0006996]	6.17E-09	41
RNA metabolic process [GO:0016070]	5.59E-08	39
cellular component organization [GO:0016043]	1.22E-07	48
regulation of gene expression [GO:0010468]	1.34E-07	35
cellular component organization or biogenesis [GO:0071840]	1.44E-07	49
nucleic acid metabolic process [GO:0090304]	3.8E-07	40
histone acetylation [GO:0016573]	1.42E-06	10
internal protein amino acid acetylation [GO:0006475]	1.82E-06	10
internal peptidyl-lysine acetylation [GO:0018393]	1.82E-06	10
peptidyl-lysine acetylation [GO:0018394]	1.82E-06	10
regulation of RNA metabolic process [GO:0051252]	1.84E-06	30
peptidyl-amino acid modification [GO:0018193]	2.04E-06	17
regulation of cellular process [GO:0050794]	2.18E-06	51
regulation of nucleobase-containing compound metabolic process [GO:0019219]	3.19E-06	30
regulation of biological process [GO:0050789]	3.53E-06	53
regulation of transcription, DNA-templated [GO:0006355]	4.64E-06	28
regulation of nucleic acid-templated transcription [GO:1903506]	4.64E-06	28
regulation of RNA biosynthetic process [GO:2001141]	4.64E-06	28
protein acetylation [GO:0006473]	5.2E-06	10
cell cycle [GO:0007049]	7.68E-06	23

Table 6 SUMOylation-independent PIAS interactors in S2 cells

Proteins with less than two log_2 fold change in PIAS IP-MS with and without NEM categorised into groups based on their biological processes. Proteins with multiple functions are listed in one category only. Underlined proteins are involved in chromosome organization. In bold are PIAS interactors validated by western blot.

Role	Proteins
Transcription regulation	**BEAF-32** , bip2, Cfp1, CG1815, dbr, enok, Hcf, hyx, koko, mip130, mip40, rept, Sec13, spen, Ssp, stwl, Taf5, Taf6, Taf8
Mitotic cell cycle	Bub1, Cdc16, **Cenp-C,** Map205, mars, Mink, Mtor, Myb, Nup43, Nup44A, pont, pzg, Spc105R
mRNA processing	Bug2, CG31368, Cpsf160, Cpsf73, cyp33, Sym, Trf4-1, hfp, l(2)37Cb, rin, Rrp40, SmD3, vir
Chromatin organisation	**ADD1,** Ibf1, Ibf2, Su(var)2-10, Su(var)3-7
Histone modification	Br140, e(y)1, Eaf6, MRG15, Rbbp5, Set1, UPset, Wdr82, wds
Molecular scaffolding	BtbVII, mxc, hts, pyd, Svil
Import into Nucleus	CG14712
Synaptic vesicle cycle	Git
Oxidative stress	Pdi
Signaling pathways	Rno, RtGEF
Circadian rhythms	tyf
Unknown function	CG1737, CG34159, CG3548, CG43736, CG4788, CG4951, CG5726, CG5787, CG9330

Table 7 Top gene ontology (GO) terms associated with proteins identified in PIAS IP-MS from S2 cells, in NEM conditions only

The analysis is a PANTHER overrepresentation test against the complete A) GO-Slim Biological Process B) GO-Slim Molecular Function C) Cellular Component annotation datasets and D) Reactome pathways with Bonferroni correction applied. N=731

(A) PANTHER GO-Slim Biological Process	Fold Enrichment	P value
mitotic sister chromatid segregation	9.82	0.023
mRNA export from nucleus	9.42	0.029
mRNA-containing ribonucleoprotein complex export from nucleus	9.42	0.029
ribonucleoprotein complex localization	8.73	0.044
ribonucleoprotein complex export from nucleus	8.73	0.044
DNA-dependent DNA replication	8.01	0.002
rRNA processing	7.57	0.009
RNA catabolic process	6.12	0.016
DNA replication	5.85	0.003
cellular macromolecule biosynthetic process	5.61	0.005

(B) PANTHER GO-Slim Molecular Function	Fold Enrichment	P value
DNA-directed 5'-3' RNA polymerase activity	7.36	0.0451
5'-3' RNA polymerase activity	7.36	0.0451
RNA polymerase activity	7.36	0.0451
catalytic activity, acting on RNA	3.88	0.00793
ligase activity	3.51	0.00141
RNA binding	3.43	5.26E-08
nucleic acid binding	2.76	2.3E-11
heterocyclic compound binding	2.72	3.05E-11
transferase activity, transferring phosphorus-containing groups	2.42	0.0202
transferase activity	1.96	0.00105

(C) PANTHER GO-Slim Cellular Component	Fold Enrichment	P value
preribosome, large subunit precursor	13.77	0.0000481
90S preribosome	10.86	0.0000595
preribosome	10.43	4.62E-12
proteasome accessory complex	10.24	0.00598
proteasome regulatory particle	10.24	0.00598
t-UTP complex	9.31	0.00000395
nucleolus	9.26	2.07E-17
small-subunit processome	8.91	0.000989
nucleolar part	8.1	0.0000161
nuclear chromosome	5.85	0.0481

(D) Reactome pathways	Fold Enrichment	P value
Processive synthesis on the lagging strand	16.83	0.01
Lagging Strand Synthesis	11.88	0.0483
Activation of the pre-replicative complex	10.1	0.00163
DNA Replication Pre-Initiation	9.77	0.00205
M/G1 Transition	9.77	0.00205
Degradation of CLK	7.48	0.0417
RNA polymerase II transcribes snRNA genes	5.52	0.0448
Synthesis of DNA	5.24	0.0000298
DNA Replication	5.04	0.0000529
G1/S Transition	5.03	0.000632

Table 8 Top gene ontology (GO) terms associated with proteins identified in PIAS IP-MS from S2 cells, in the absence of NEM

The analysis is a PANTHER overrepresentation test against the complete A) GO-Slim Biological Process B) GO-Slim Molecular Function C) Cellular Component annotation datasets and D) Reactome pathways with Bonferroni correction applied. N=134

(A) PANTHER GO-Slim Biological Process	Fold Enrichment	P value
gene expression	3.34	0.0305
organic substance metabolic process	2.42	0.0275

(B) PANTHER GO-Slim Molecular Function	Fold Enrichment	P value
DNA binding	5.04	0.0121

(C) PANTHER GO-Slim Cellular Component	Fold Enrichment	P value
nuclear part	4.16	0.0464
nucleus	2.85	0.0352

(D) Reactome pathways	Fold Enrichment	P value
RNA Polymerase II Pre-transcription Events	15.93	0.0038
RNA Polymerase II Transcription	5.05	0.0376

4.3.10 Validation of PIAS interactors by co-immunoprecipitation followed by western blot

For validation of endogenous PIAS interactions identified by PIAS IP-MS, IP followed by western blot was performed for targets of interest where antibodies were available. As already shown, HP1a (Figure 4–4) and XNP (Figure 4–13) co-immunoprecipitate with PIAS. PIAS interacts with ADD1 in the absence or presence of NEM. BEAF-32 and Cenp-C followed the same pattern (Table 6).

PIAS interaction with ADD1, BEAF-32, Cenp-C and HP1a was validated by immunoprecipitation of PIAS followed by blotting with antibodies specific to the predicted interaction partners (Figure 4–18). The PIAS-ADD1 interaction is maintained in +/-NEM conditions (Figure 4–18A), suggesting that it is not strongly influenced by SUMOylation. BEAF-32 co-immunoprecipitated with PIAS (Figure 4–18Bi). However, in the reverse experiment PIAS did not co-immunoprecipitate with BEAF-32 (Figure 4–18Bii). This may be because the BEAF-32 antibody prevents interaction between the two proteins (e.g. by recognising an epitope that is within their binding site), therefore preventing the co-immunoprecipitation of PIAS-BEAF-32 complexes. BEAF-32 did not co-immunoprecipitate with SUMOylated proteins and SUMOylation of BEAF32 is not obvious on western blot (data not shown). This supports that PIAS interacts with BEAF-32 in a manner independent of SUMOylation. PIAS and Cenp-C co-immunoprecipitated (Figure 4–18C), but whether this is a soluble pool of Cenp-C or one capable of integrating into chromatin would require further characterisation.

HP1a was only slightly enriched in PIAS IP-MS in NEM conditions (Figure 4–17), therefore HP1a was probed for in PIAS IP. HP1a co-immunoprecipitates with PIAS in both the absence and present of NEM, with an additional band (migrating at a slower rate) in NEM conditions (Figure 4–18D). The identify of this band was not confirmed.

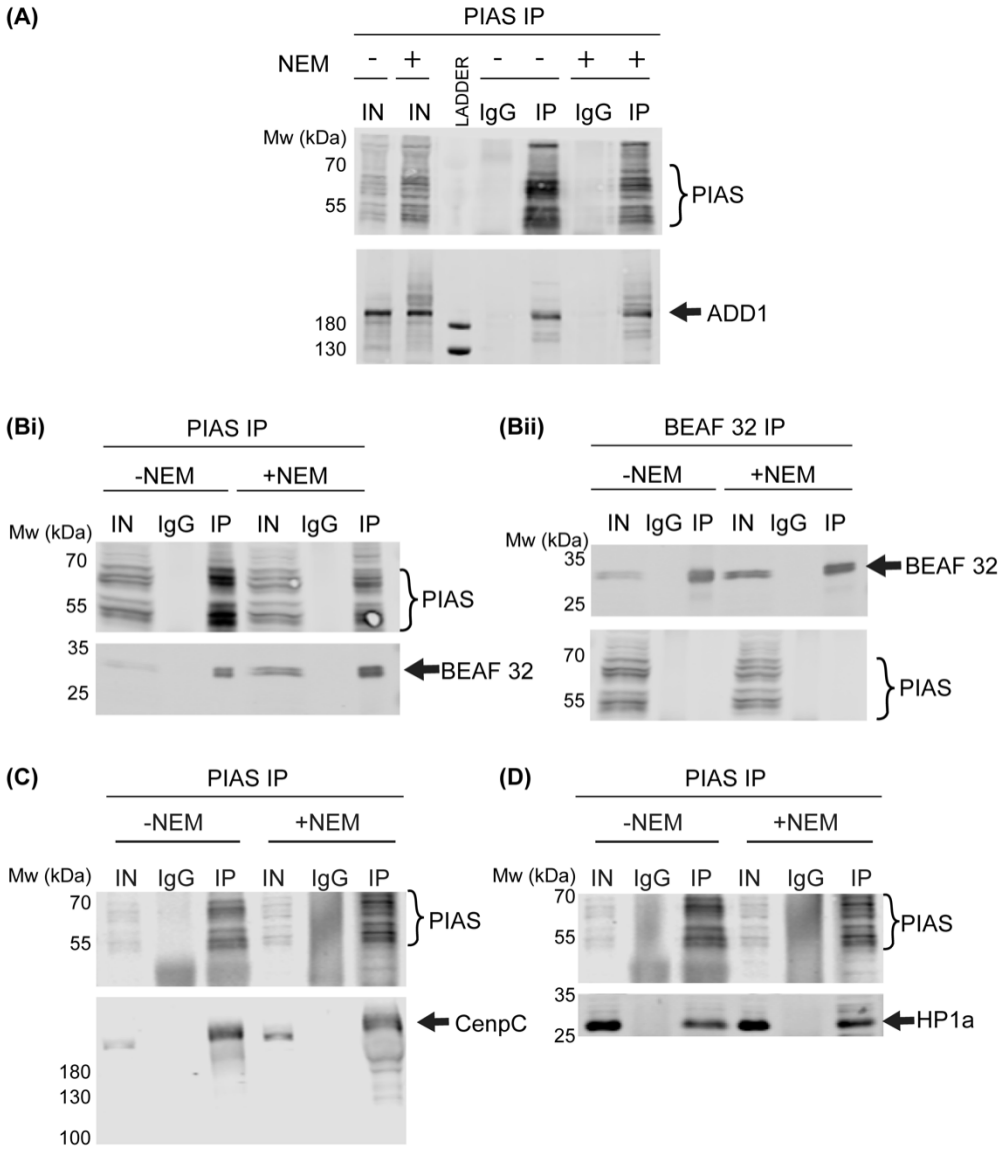


Figure 4–18 PIAS co-immunoprecipitates with ADD1, BEAF-32, Cenp-C and HP1a regardless of the presence of NEM

Nuclear extracts were made from wild-type Schneider S2 cells. The input (IN), IP with affinity purified anti-IgG sheep antibody coupled to Protein G Dynabeads™ (IgG) and IP with anti-PIAS sheep antibody coupled to Protein G Dynabeads™ (IP) (PIAS IP) were analysed on 4-20% SDS-PAGE followed by western blot probed with *Drosophila* anti-PIAS sheep antibody and anti-ADD1 rabbit antibody **(A)** or anti-BEAF-32 mouse antibody **(B)** or anti-Cenp-C rat antibody **(C)** or anti-HP1a mouse antibody **(D)**.

(Bii) IP with affinity purified anti-IgG mouse antibody coupled to Protein A Dynabeads™ (IgG) and IP with anti-BEAF-32 mouse antibody coupled to Protein A Dynabeads™ (IP)

4.3.11 PIAS post-translational modification identification

Post-translational modifications (PTMs) of PIAS have been previously identified. These may influence its interactions with other proteins. Mass spectrometry identified peptides mapping to PIAS that were modified by phosphoryl groups, acetyl groups and SMT3, which indicates SUMOylation. These sites are marked on the schematic diagram of PIAS isoforms (Figure 4–19). A SUMOylation site was found in exon 4 of PIAS in the RLD of the PIAS-CORE region. Phosphorylation was found in exon 9, an exon found only in isoforms J and L, which could be important for cell cycle regulation. These peptides were identified from PIAS IP-MS from S2 cells by one peptide in one out of six experiments, suggesting that these are dynamic modifications, present on few proteins in the cell.

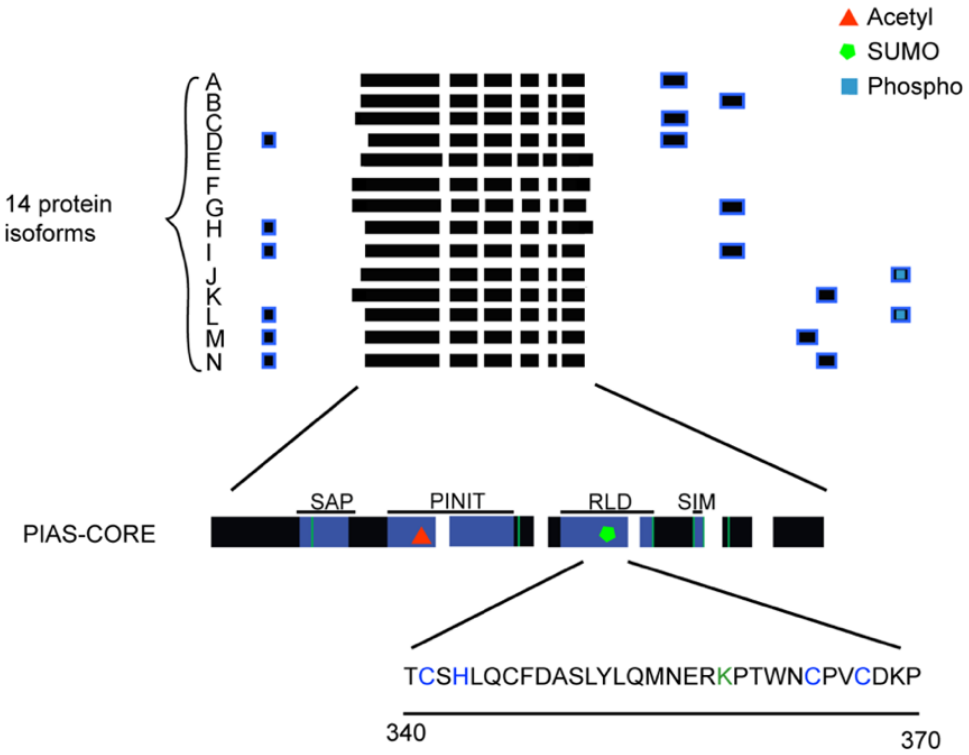


Figure 4–19 Post-translational modifications on PIAS

Phosphorylated peptides were found mapping to the C-terminus exons in isoforms J and L, Acetylation was found mapping to exon 2 in PIAS-CORE region and SUMOylation was found in the Ring-like domain (RLD). The amino acid sequence of PIAS RLD is shown, residues in blue mark the amino acid residues important for the RING structure. K where SUMO is conjugated is marked in green.

4.3.12 Generation of PIAS^{V5} CRISPR Cas9 tagged cell lines

Single guide RNAs (sgRNA) and homology directed repair templates (HDRT) were designed to target the core region of the *su(var) 2-10* locus on chromosome 2R using CRISPR/Cas9 in S2 cells (Figure 4–20A). Single guide RNAs targeting the N-terminus of the PIAS core region at the *Su(var) 2-10* gene locus resulted in reduced expression of the higher molecular weight cluster of PIAS isoforms (Figure 4–20B). Only the higher isoforms incorporated the V5 tag (Figure 4–20C). No PIAS-N-V5 S2 cell line clones with all copies of PIAS tagged were recovered.

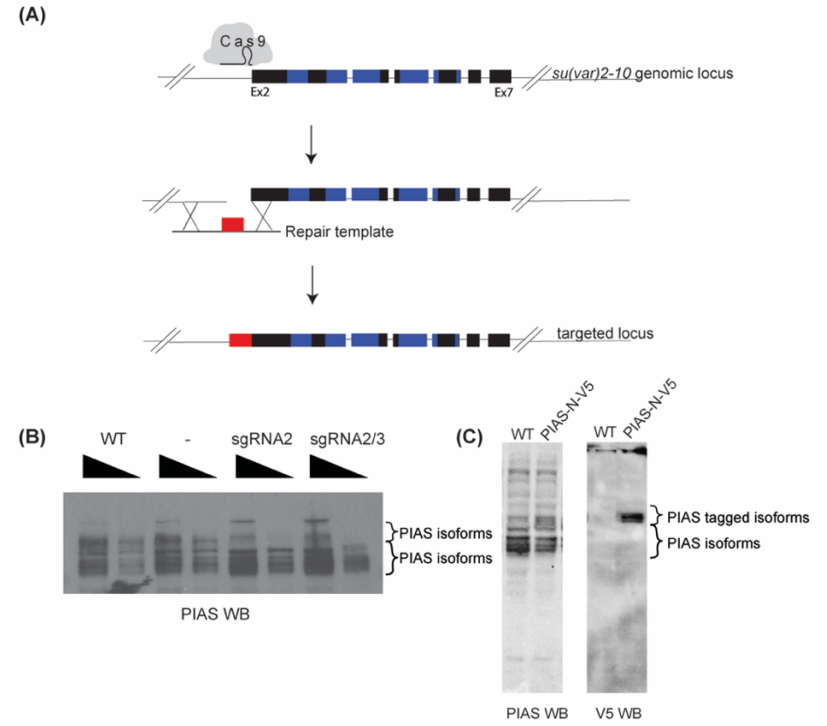


Figure 4–20 Generation of the PIAS-N-V5 allele

(A) Schematic diagram of the targeting strategy using CRISPR/Cas9 for the insertion of a V5 tag at the N-terminus of PIAS- CORE by targeting the *Su(var)2-10* locus on chromosome 2R in *Drosophila* S2 cells to create the cell line PIAS-N-V5. **(B)** sgRNAs targeting *Su(var)2-10* to test sgRNA efficiency through knock-down of PIAS levels. WT=wild-type cell line without transfection, - control sgRNAs, sgRNA2 and sgRNA 3 target the *Su(var)2-10* gene locus. Equal concentrations of protein were analysed by SDS-PAGE followed by western blot probed with *Drosophila* anti-PIAS sheep antibody **(C)** Western blot analysis of protein extracts from wild-type (WT) and PIAS-N-V5 cell lines, left-hand-blot probed with anti-PIAS sheep antibody recognising endogenous protein and right-hand-blot probed with anti-V5 mouse antibody recognising tagged isoforms.

The L2-4 clone of S2 cells used in the Heun lab are pseudo-tetraploid with 13 chromosomes (2x X, 4x 2nd, 4x 3rd, 2x 4th, 1x 2L fragment). Therefore, for all PIAS proteins to be tagged, all four *Su(var)2-10* alleles on chromosome 2R must be targeted.

With C-terminal tagging of PIAS- CORE, it was possible to derive knock-in S2 cell lines expressing a V5-tag at the C-terminus of PIAS (PIAS-C-V5) from all four copies of the endogenous *Su(var)2-10 l*ocus (describe in Figure 4–21), resulting in all endogenous PIAS proteins being tagged. Integration of the tag into the genomic DNA was confirmed by PCR and cell lines were plated at low density to form clonal populations which were screened for cells in which all four alleles of *Su(var)2-10* were successfully tagged, resulting in the cell line PIAS-C-V5 (Figure 4–21C). PIAS-C-V5 was validated by western blot (Figure 4–21D). In the process of incorporating the tag, some isoforms of PIAS were lost, potentially by tagging affecting alternative splicing. Cells doubled at a normal rate and maintained the tag, suggesting that the V5 tag did not impair PIAS function. Immunofluorescence of the V5 tag in PIAS-C-V5 showed a similar general nuclear localisation pattern as observed for endogenous PIAS in interphase S2 cells (Figure 4–21E). While PIAS-CORE overexpression from pMT-Pias9_V5 showed localisation in nuclear bodies (Figure 4–7), PIAS bodies were not observed using endogenously tagged cell lines or by staining with the endogenous antibody (Figure 4–21E) at any stage in the cell cycle. PIAS-C-V5 was also amenable to immunoprecipitation using the *Drosophila* anti-PIAS sheep antibody (Figure 4–21F). In wild-type cells the anti-PIAS sheep antibody recognises all isoforms of PIAS whereas the number of isoforms of PIAS in PIAS-C-V5 is reduced (Figure 4–21F).

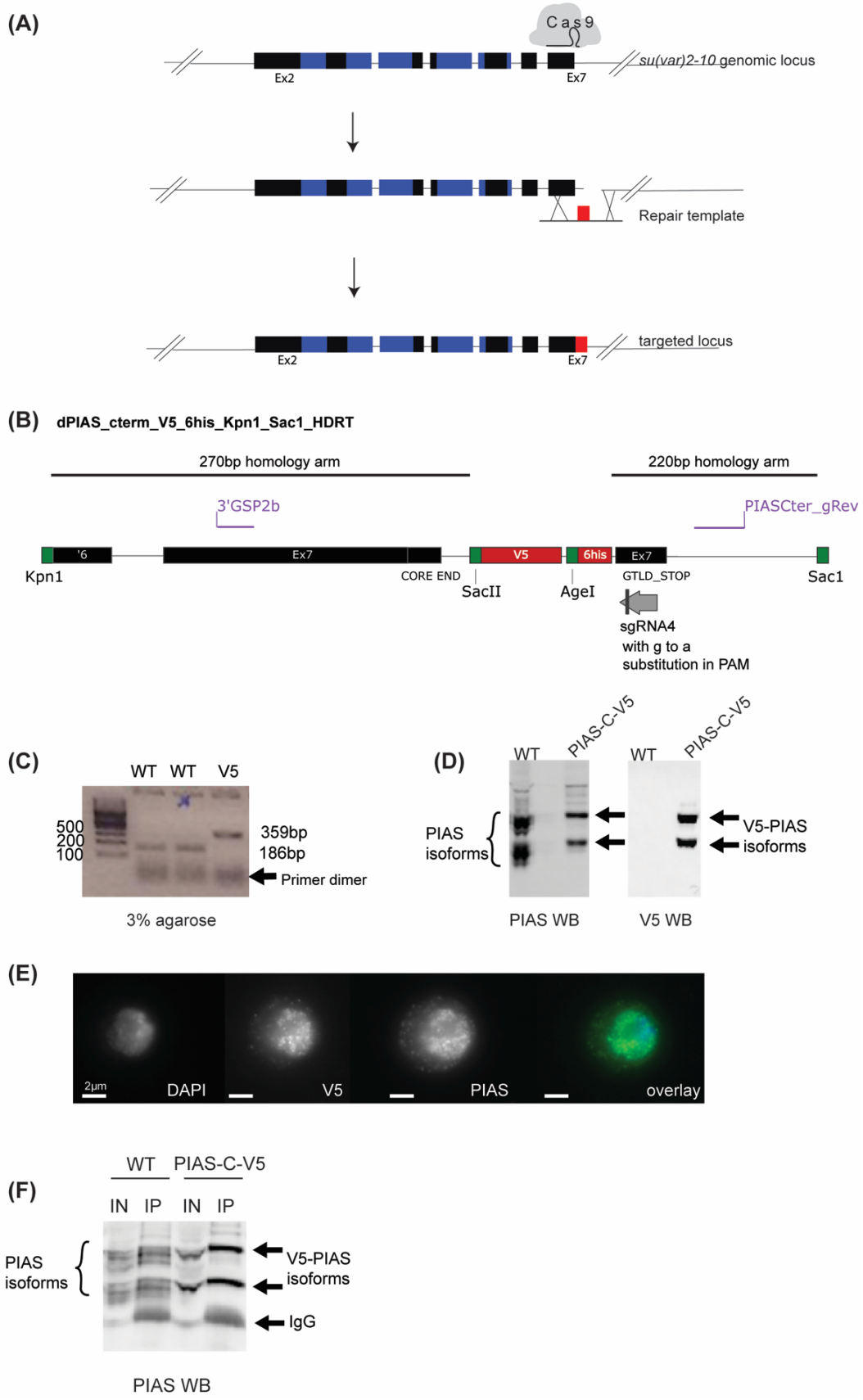

(A)
Cas9
su(var)2-10 genomic locus
Ex2
Ex7
Repair template
targeted locus
Ex2
Ex7
(B)
dPIAS_cterm_V5_6his_Kpn1_Sac1_HDRT
270bp homology arm
220bp homology arm
3'GSP2b
PIASCter_gRev
'6
Ex7
V5
6his
Ex7
Kpn1
CORE END
SacII
AgeI
GTLD_STOP
Sac1
sgRNA4
with g to a
substitution in PAM
(C)
WT WT V5
500
200
100
359bp
186bp
Primer dimer
3% agarose
(D)
WT
PIAS-C-V5
WT
PIAS-C-V5
PIAS
isoforms
V5-PIAS
isoforms
PIAS WB
V5 WB
(E)
2µm
DAPI
V5
PIAS
overlay
(F)
WT
PIAS-C-V5
IN IP IN IP
PIAS
isoforms
V5-PIAS
isoforms
IgG
PIAS WB

Figure 4–21 Generation of the PIAS-C-V5 allele

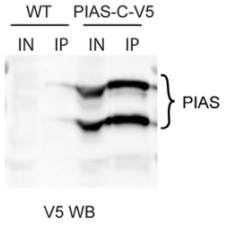
(A) Schematic diagram of the targeting strategy using CRISPR/Cas9 for the insertion of a V5 tag at the C-terminus of the PIAS core region on chromosome 2R in *Drosophila* S2 cells to create PIAS-C-V5. **(B)** Schematic diagram of the repair construct showing the recognition site of the sgRNA used for targeting and the primers used to confirm insertion of the tag at the endogenous locus. The homology arms are not to scale. **(C)** 3% agarose gel analysis of PCR products, showing a band shift from 186 to 359bp when the primers amplify across the locus containing the inserted repair construct **(D)** Western blot of protein extracts from wild-type and PIAS-C-V5 cell lines, left-hand blot probed with anti-PIAS sheep antibody recognising endogenous protein and right hand blot anti-V5 antibody recognising tagged isoforms. **(E)** Immunofluorescence staining of untagged and V5-tagged PIAS proteins in S2 cells. PIAS-C-V5 visualised by an anti- V5 antibody and untagged PIAS visualised by the anti-PIAS sheep antibody display nuclear localisation. Scale bar=2µm **(F)** Western blot of input (IN) and immunoprecipitation (IP) samples using the anti-PIAS sheep antibody from wild-type (WT) and Pias-C-V5 S2 cell lines probed with anti-PIAS sheep antibody.

4.3.13 Immunoprecipitation of PIAS-C-V5

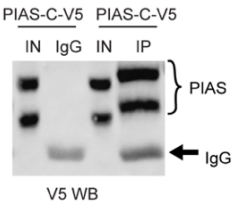
To validate endogenous PIAS IPs using the *Drosophila* anti-PIAS sheep antibody, I performed IPs using the V5 tag in PIAS-C-V5 S2 cell lines (Figure 4–22A and B). PIAS-C-V5, allows analysis of PIAS's interaction partners in the absence of any PIAS antibody-generated background. IP using the V5 tag in PIAS-C-V5 cells did not show the expected enrichment of SUMOylated proteins in +NEM conditions by western blot (Figure 4–22C) compared to IP using the endogenous antibody (Figure 4–10). It is possible that specific PIAS isoforms, lost with the V5 tag, are needed to form some PIAS-SUMO containing complexes, or that IP via the V5 tag immunoprecipitated SUMOylated proteins at below the limit of detection of western blot. However, IP-MS using the V5 tag to immunoprecipitate PIAS-C-V5 did replicate findings from the endogenous PIAS IP-MS, finding many of the same proteins including the FACT complex subunit (dre4), Topoisomerase 2 (Top2) and XNP (Figure 4–22D). IP-MS from PIAS-C-V5 was carried out once and so statistics were not performed on interaction partners identified and data is presented as a table. In addition, a greater number of peptides and proteins were identified in +NEM conditions (Figure 4–22E).

To validate V5-IP of PIAS-C-V5, ADD1 was probed for on western blot. This showed that ADD1 co-immunoprecipitated with PIAS-C-V5 (Figure 4–22F) regardless of the presence of NEM, exhibiting the same behaviour as observed with endogenous PIAS IP. Endogenous interactors of PIAS identified from IP with *Drosophila* anti-PIAS sheep antibody and V5-IP from PIAS-C-V5 in +NEM conditions were compared, identifying 519 PIAS interactors in common, including: HP1a, XNP, SMT3 and H3.3 (Figure 4–23).

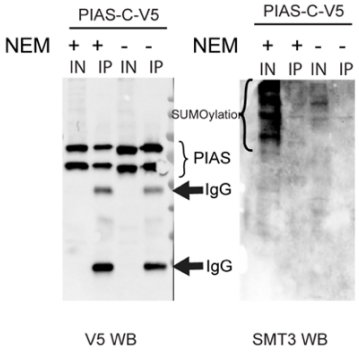
(A)



(B)



(C)



(D)

Majority Protein ID	Gene Name	WT -NEM	WT +NEM	PIAS-C-V5 -NEM	PIAS-C-V5 +NEM
P15348	**Top2**	6	6	29	100
A1Z8U0	Prp8		11	15	97
Q59E58	zip				86
Q99323-2	zip				85
Q9VUV9	l(3)72Ab		11	14	82
Q8IRG6	**dre4**	1	4	12	62
Q9VC94	CG5728				50
Q8MSS0	CG7839				46
Q9W0B8	alphaCOP			8	46
M9PIA6	**Mi-2**				45
Q8INM3	sle			5	44
X2JG50	CG14215			1	44
Q8IH00	Mat89Ba			1	42
Q9VM75	l(2)k09022				41
D0Z769	RpII215-RA			6	41
Q9VAX8	CG4849		8	8	40
Q9VK59	CG5787			13	40
Q9VCW3	Nup133				37
B5RJ65	**Ssrp-RA**				36
M9PBH8	CG2199	1	1	1	36
P13060	EF2			13	35
Q24368	**Iswi**				35
A0A0B4K7Y7	larp				35
E1JGV6	Nop60B	6		7	35
A0A0B4LI86	RpII140			6	34
F3YDH0	Hsc70-3	4	3	13	34
A1YK62	Nup107				33
A1Z909	CG8545-RA			4	33
Q9VUE5	**stwl**				33
Q9VID9	CG8677				32
Q9W0R0	Cdc5		1	4	32
X2JEW2	CG2691				32
O16797	RpL3	2	2	12	31
P09180	RpL4	5	7	17	31
C7LA75	Hsc70-4		4	7	31
M9MRS7	Myo31DF			1	31
Q7K0X3	CG2469				31
Q9GQV2	cher			32	31
Q9VEX6	bor				31
Q9VUR2	CG12301			3	31
Q9VVU6	CG6841		2	5	31
A1ZA92	CG8414				30
A0A0B4LI25	ncd				30
Q9VD07	CG12499				30
X2J9F3	Nup160				30
H0RNM5	l(1)G0020-RA				29
Q9VHC7	rump			6	29
Q9W5E4	l(1)1Bi				29
A1Z6Z9	didum				28
Q9VT61	CG8108			2	28
A0A0B4LHR3	**XNP**				27
M9PCC1	Rack1	3	5	7	27
Q7JQN4	Rs1				27
Q9VGG9	CG31368				27
X2JEU3	CG1737				27
P13395	alpha-Spec				26
G3M3A2	RpS3	5	6	14	26
M9NDX3	Top1		5	3	26
Q9VIG1	CG12050-RB				26
P55830	RpS3A			9	25
Q8T6B9	pUf68		5		25
A0A0B4K7G4	mod			4	25
Q7JXU4	Ns2				25
Q9U5W4	nop5		2	4	25

(E)

	WT-	WT+	PIAS-C-V5-	PIAS-C-V5+
no. peptides	229	357	1059	8101
no.proteins	66	119	256	781

(F)

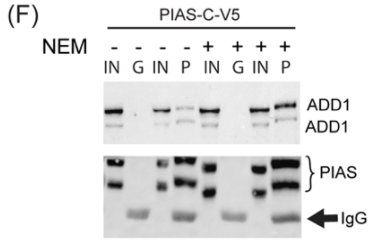
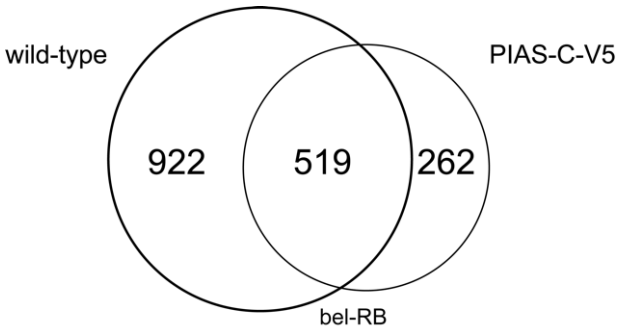


Figure 4–22 PIAS-C-V5 interactors

(A) Western blot showing input (IN) and immunoprecipitation (IP) of V5 tagged species from wild-type (WT) or PIAS-C-V5 Schneider S2 cell nuclear extract, probed with mouse anti-V5 antibody. **(B)** Immunoprecipitation of PIAS using the *Drosophila* anti-PIAS sheep antibody or immunoprecipitation using sheep anti-IgG sheep antibody as a control from PIAS-C-V5 Schneider S2 cell nuclear extract, western blot with anti-V5 mouse antibody. **(C)** Immunoprecipitation of PIAS using the *Drosophila* anti-PIAS sheep antibody from PIAS-C-V5 Schneider S2 cell nuclear extract in the presence and absence of NEM. Samples were run on 4-20% SDS-PAGE for western blot probed with *Drosophila* anti-PIAS sheep antibody and the gel was re-probed with anti-SMT3 rabbit antibody. **(D)** Table: List of Proteins identified in a V5-IP from PIAS-C-V5 Schneider S2 cell nuclear extract using anti-V5 agarose with WT cells as a negative control. **(E)** Table listing the number of peptides and proteins in PIAS-C-V5 mass spectrometry. **(F)** Immunoprecipitation of PIAS using the *Drosophila* anti-PIAS sheep antibody (P) or immunoprecipitation using anti-IgG sheep antibody (G) from PIAS-C-V5 Schneider S2 cell nuclear extract in the presence and absence of NEM. Input (IN) and IP samples were run on 4-20% SDS-PAGE for western blot probed with anti-PIAS sheep antibody or anti-ADD1 rabbit antibody.



bel-RB
Cenp-C
Dcr-2-RA
dre4
His3.3A
Ibf2
Iswi
JIL-1
Mi-2
Rm62
row
Set1
SMT3
snRNP-U1-70K
Ssrp-RA
Su(var)2-HP2
HP1a
Su(var)2-10
Su(var)3-7
Ulp1
XNP

Figure 4–23 PIAS interactors in +NEM conditions

Venn diagram of proteins identified in PIAS IP-MS from wild-type and PIAS-C-V5 S2 cells. Chromatin associated proteins of interest found in both conditions listed in alphabetical order. Those discussed in this thesis are in bold.

4.4 Summary

This chapter described the method development for identification of endogenous PIAS interaction partners by immunoprecipitation followed by mass spectrometry (IP-MS) with the aim to define interaction partners of PIAS. PIAS was immunoprecipitated from S2 cell lysate in the presence and absence of NEM, to identify interactions that are maintained in the presence and absence of SUMOylation. PIAS IPs were performed using the *Drosophila* anti-PIAS sheep antibody and a tagged version of PIAS proteins to cross-validate the PIAS interactors identified. PIAS IP-MS identified many PIAS interactors, including those involved in chromosome organization, transcription, mRNA processing (Table 5) and RNA interference pathways (Figure 4–15 and Figure 4–16).

Characterisation of PIAS in NEM conditions, where SMT3 conjugation to substrate proteins is maintained (compared to when NEM is not present), clarified which PIAS interactors depend on the presence of SUMOylation. Overexpression of PIAS in S2 cells causes sequestration into nuclear bodies probably dependent on large SIM-SUMO networks (Figure 4–7).

XNP co-immunoprecipitated with PIAS only in the presence of NEM (Figure 4–13), supportive of the hypothesis that PIAS SUMOylates XNP. However, by PIAS IP-MS, XNP was not identified as a top PIAS interactor in wild-type S2 cells (Figure 4–15 and Figure 4–16), this could be a limitation of using mass spectrometry (which requires the presence of unique peptides to identify the protein and requires that these peptides fly in the experiment) and should be confirmed by alternative methods. However, size exclusion chromatography separating PIAS-complexes suggests the presence of many complexes which don't contain XNP (Figure 4–2).

I have confirmed an interaction between PIAS and heterochromatin protein HP1a (Figure 4–4), validating an interactor from the HP1a interactome of Alekseyenko *et al.,* 2014[299] (Table 4). I also show by size exclusion chromatography that PIAS and HP1 proteins co-elute and co-immunoprecipitate (Figure 4–5), suggesting the identity of different HP1a-PIAS complexes.

4.4.1 Evaluation of the approach

4.4.1.1 Immunoprecipitation using endogenous antibodies

As SUMOylated proteins can form large complexes through SUMO-SIM interactions, expression levels are important for understanding physiological interactions of SUMOylated proteins. Previous work has demonstrated that PIAS levels have to be maintained within a certain level; as both over expression and knockdown have lethal effects *in vivo*[144]. PIAS proteins have multiple functions in *Drosophila,* one of which is to enhance SMT3 conjugation to substrate proteins through their E3 ligase activity. NEM was used as an inhibitor of de-SUMOylation in lysis buffers to prevent the action of SUMO proteases. This meant that there was no disruption to cellular metabolism by addition of enzymes or knockdown of proteins prior to cell lysis and endogenous protein levels could be assessed.

Many studies have focused on the identification of SMT3 and SUMO substrates by proteomic approaches using mass spectrometry[199]. These studies vary in their organism, tissue or cell type of choice, combined with using different stresses, buffers and growth conditions which give maximum conjugation of SMT3 or SUMO to target proteins. Two main types of experiments have been performed: looking at substrates covalently modified by SUMO using denaturing conditions, or less stringent IPs, under non-denaturing conditions, where large SUMO-SIM networks have been maintained. Despite SMT3-substrates being identified and the SUMOylation of those substrates being characterised, the enzymes responsible for targeting SMT3 conjugation have not been so well characterised.

Since NEM is a general inhibitor of cysteine proteases, it prevents the action of any enzyme that contains thiol residues in its active site, including all UBLs. Therefore, further work is required to understand whether a SUMOylation-dependent function of PIAS is part of its mechanism of action as a chromatin factor. Although there are caveats associated with the use of NEM, it has the advantage of not interfering with cellular processes, only taking effect post-lysis to preserve interactions. An alternative experiment to identify PIAS interactors that depend on PIAS's SUMO E3 ligase activity, would be to compare endogenous PIAS IP from wild-type and the homozygous *Su(var)2-10*^{1} *Drosophila* line, which contains a point mutation in the

catalytic RING domain, inactivating PIAS's E3 ligase activity. However, there are no available data describing this mutant in homozygosity. In addition, it is likely to be lethal. When trans-heterozygous with *Su(var)2-10*^{2}, aberrant chromosome defects make amplification of *Drosophila* difficult, creating a challenge for biochemical experiments that require a large amount of starting material. A knock-in strategy using CRISPR/Cas9 to introduce a inducible point mutation in the *Su(var)2-10* locus, as has recently been performed to explore the catalytic activity of Polycomb repressive complex 2[315], would be the optimal strategy. However, this would require editing of all four alleles in S2 cells, would be technically challenging and may be lethal once induced, when E3 ligase activity is lost.

4.4.1.2 IP-MS data analysis to compare interactors in the presence and absence of NEM

Scatter plots showing $\log_2$ transformed LFQ data allowed separation of three populations of PIAS interactors (Figure 4–17): those that only interact with PIAS in the absence of NEM; those that only interact with PIAS in the presence of NEM; and those that interact with PIAS in both conditions.

Volcano plots comparing IP-MS in conditions with and without NEM fail to identify proteins that interact with PIAS independently of SUMOylation to those that significantly differ in binding with and without NEM (found both with and without NEM). Therefore scatter plots of protein abundance (with IgG control subtracted) with and without the presence of NEM illustrates that there was a huge increase in protein binding in the presence of NEM (Figure 4–17), indicating that a number of interactors are only observed when SUMOylation is maintained (XNP and Dcr-2).

4.4.1.3 Generation of tagged cell lines

Affinity purification using endogenous antibodies has the disadvantage of needing a highly specific and efficient antibody against the target protein and requires a suitable control for background levels of non-specific immunoprecipitation. Immunoprecipitation of PIAS by a V5 epitope-tag independently validates interactors identified with the *Drosophila* anti-PIAS sheep antibody (Figure 4–23).

Tagging of the core region of PIAS resulted in disruption of isoforms (Figure 4–21); the extent to which this affects PIAS function in S2 cells is unknown as no obvious defects were observed. It is thought that the PIAS isoforms may play important roles during development due to partial rescue of lethality in trans-heterozygous *Su(var)2-10*^{1}*/Su(var)2-10*^{2} mutants by expression of PIAS-PE[144]. An alternative approach would be to delete the endogenous locus by insertion of *LoxP* sites and expression of a *Cre* recombinase alongside insertion of the cDNA sequence (without introns) to replace the *Su(var)2-10* locus to overcome the problem of alternative splicing interference and multiple isoforms of PIAS. Also, by amplifying different cDNAs, it would be possible to address which PIAS isoforms are necessary for certain interactions. However, the health of the cells may be compromised, as it has not been demonstrated in a precise PIAS knockout that the CORE region is sufficient to rescue PIAS knockdown. Isoform-specific interaction studies or studies in a system with fewer PIAS isoforms could help decipher the action and subcellular distribution of PIAS proteins and determine their importance at heterochromatin and within different PIAS complexes *in vivo*.

Chapter 5 Identification of PIAS interactors in *Drosophila* embryos

5.1 Objectives

As the Heun lab was interested in the role of PIAS in the establishment of heterochromatin domains through SUMOylation of chromatin proteins, and in light of recent data linking PIAS to the Piwi/piRNA pathway[126,316], I decided to perform PIAS IP-MS from *Drosophila* embryos at different timepoints in *Drosophila* development with the aim of identifying proteins responsible for heterochromatin formation that could possibly require PIAS mediated SUMOylation for their function.

The aim of this chapter was to identify PIAS interactors in *Drosophila* embryos. I used the previously tested affinity purified *Drosophila* anti-PIAS sheep antibody that recognises all PIAS isoforms (Chapter 4) for immunoprecipitation of PIAS from wild-type embryos. Firstly, I compared mixed populations of embryos either in nuclear cycles 1-14 (1-3hours) during the blastula stages of embryogenesis or after the nuclear cycles were complete (12-15 hours) post-gastrulation once tissues had been specified. I went onto identify PIAS interactors at more specific timepoints during the mid-blastula transition and heterochromatinization. Finally, I used the data from S2 cells (presented in Chapter 4) to complete a bioinformatic comparison of PIAS interactors identified in S2 cells versus embryos at different timepoints in development to identify interactors specific to early *Drosophila* embryos. In the Schneider 2 (S2) cell line (clone L2-4 originally derived from 20-24hr *Drosophila melanogaster* embryos), PIAS does not localise to heterochromatin and is instead found at euchromatic regions[57]. Therefore, comparison of PIAS IP-MS from timed *Drosophila* embryos and S2 cells could allow categorisation of constitutive PIAS interactors (found to co-immunoprecipitate with PIAS in all conditions) or identify proteins that interacted with PIAS specifically in the early embryo, when PIAS is localised at heterochromatin.

5.2 Aims

- To identify PIAS interactors in *Drosophila* embryos
- To identify whether PIAS interactors change during *Drosophila* embryo development
- To identify whether ADD1 co-localises with PIAS in *Drosophila* embryos
- To identify common PIAS interactors in S2 cells and embryos
- To identify common properties, biological roles and molecular functions of PIAS interactors

5.3 Results

5.3.1 Identifying PIAS interactors *in vivo*, in *Drosophila* embryos

By identifying PIAS interactors *in vivo*, I hoped to shed light on the biological function of PIAS in early embryos where its role remains largely elusive. I started by using IP-MS to identify PIAS interactors in *Drosophila* embryos at early and late timepoints in development. Protein identified to interact with PIAS in *Drosophila* embryos between 1-3 hr (Figure 5–1) and 12-15 hr (Figure 5–2) were compared to IgG IP-MS from equivalent extracts using volcano plots to visualise proteins that co-immunoprecipitated with PIAS and were enriched in PIAS IP-MS compared to IgG control IP-MS. Proteins identified by two or more peptides and 30 or more peptides were plotted (Figure 5–1 and Figure 5–2) to emphasis high confidence PIAS interactors.

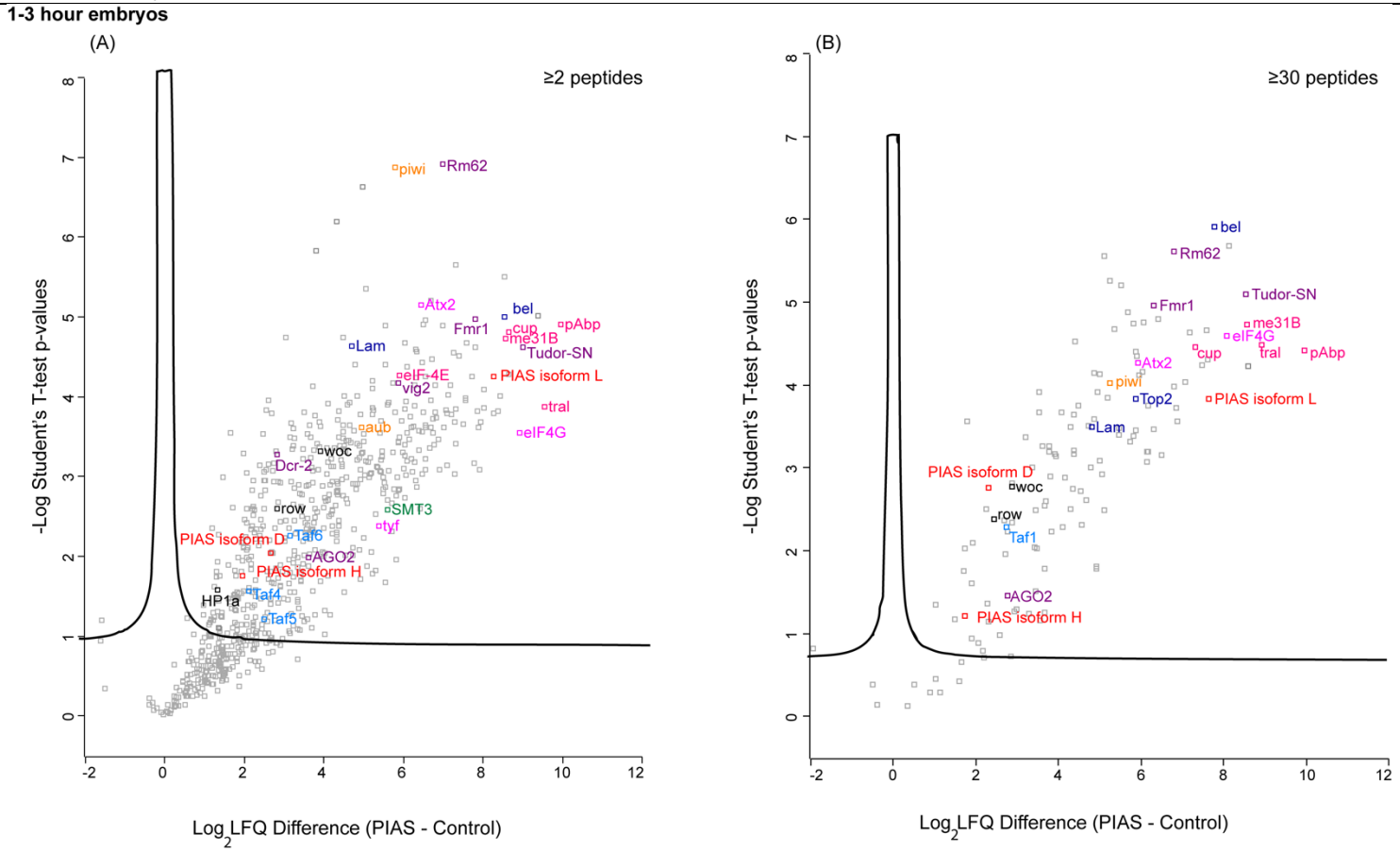

1-3 hour embryos
(A)
≥2 peptides
piwi
Rm62
Atx2
bel
Fmr1
cup
pAbp
me31B
Lam
Tudor-SN
eIF-4E
PIAS isoform L
vig2
tral
aub
eIF4G
woc
Dcr-2
row
SMT3
tyf
PIAS isoform D
Taf6
AGO2
PIAS isoform H
HP1a
Taf4
Taf5
-Log Student's T-test p-values
Log2LFQ Difference (PIAS - Control)
(B)
≥30 peptides
bel
Rm62
Tudor-SN
Fmr1
me31B
eIF4G
cup
tral
pAbp
Atx2
piwi
Top2
PIAS isoform L
Lam
PIAS isoform D
woc
row
Taf1
AGO2
PIAS isoform H
-Log Student's T-test p-values
Log2LFQ Difference (PIAS - Control)

Figure 5–1 Volcano plot of PIAS protein interactors in *Drosophila* embryos between 1-3 hours in development

Immunoprecipitation (IP) using *Drosophila* affinity purified anti-PIAS antibody (PIAS) or control sheep anti-IgG (Control) coupled to Protein G Dynabeads™ was performed from embryo lysate. IP samples were prepared as described in methods and analysed by mass spectrometry. PIAS interactors in conditions with NEM are plotted. To determine the PIAS interactome, a MaxQuant label-free quantitation pipeline was used. Student's T-test was performed on $\log_2$ transformed LFQ values from three control and three PIAS IPs after filtering for proteins identified by two or more peptides **(A)** or 30 or more peptides **(B)** and found in at least two of three replicates. T-test difference ratios (x-axis) were plotted against the negative logarithmic p-value of the T-test (y-axis). Each dot represents one protein. PIAS is in red. Proteins in purple are linked to RNAi silencing pathways. Proteins in black are consistent PIAS interactors. Proteins in navy blue are chromatin proteins. SMT3 is in green. Polyadenylate-binding protein and proteins involved in post-transcriptional regulation are in pink. The confidence line was generated using a permutation-based FDR value of 0.05 and a minimal fold change value of 0.1.

12-15 hour embryos

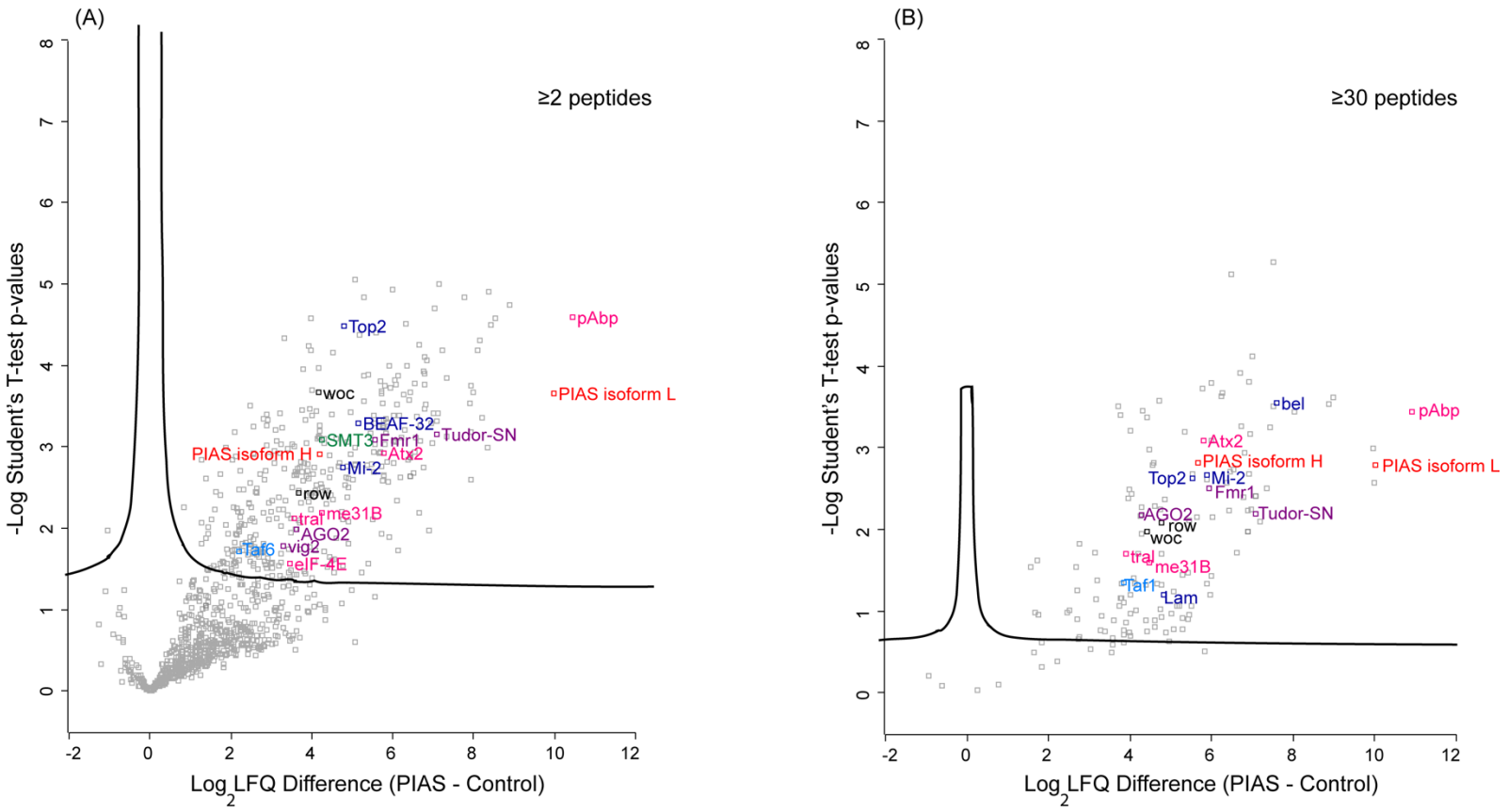


Figure 5–2 Volcano plots of PIAS protein interactors in *Drosophila* embryos between 12-15 hours in development

Immunoprecipitation (IP) using *Drosophila* affinity purified anti-PIAS antibody (PIAS) or control sheep anti-IgG antibody (Control) coupled to Protein G Dynabeads™ was performed from embryo lysate. IP samples were prepared as described in methods and analysed by mass spectrometry. PIAS interactors in conditions with NEM are plotted. To determine the PIAS interactome, a MaxQuant label-free quantitation pipeline was used. Student's T-test was performed on log_2 transformed LFQ values from three control and three PIAS IPs after filtering for proteins identified by two or more peptides **(A)** or 30 or more peptides **(B)** and found in at least two of three replicates. T-test difference ratios (x-axis) were plotted against the negative logarithmic p-value of the T-test (y-axis). Each dot represents one protein. PIAS in red. Proteins in purple are linked to RNAi silencing pathways. Proteins in black are consistent PIAS interactors. Proteins in navy blue are chromatin proteins. SMT3 is in green. Polyadenylate-binding protein and proteins involved in post-transcriptional regulation are in pink. The confidence line was generated using a permutation-based FDR value of 0.05 and a minimal fold change value of 0.1.

In 1-3hr embryos, Piwi and aub (components of the Piwi-piRNA pathway) were identified, as were other components associated with RNAi silencing complexes including: Fmr1, Tudor-SN, AGO2, vig2, bel and Dcr-2 (Figure 5–1). RISC complex components found to interact with PIAS in early embryos were also present in PIAS IP-MS from 12-15hr embryos (Figure 5–1 and Figure 5–2). Piwi was identified by more than 30 peptides in 1-3hr embryos (Figure 5–1); this is in agreement with the recent findings of Ninova *et al.,* 2019[163], who identified an interaction between PIAS and Piwi.

Other proteins active in gene silencing pathways were identified in PIAS IP-MS. PIAS interaction with pABP and proteins involved in post-transcriptional gene silencing were identified in both 1-3hr and 12-15hr embryos (Figure 5–1, Figure 5–2). These include Atx2, me31B, pcm, Upf1 and Rm62. Seven proteins found in the nuage/p-granule/germ plasm were identified: tral, TER94, Fs(2)ket, me31B, Fmr1, CG9425 and bel. Tral and cup were also identified. These interact with me31b to regulate the translation state of eIF4E1 bound mRNAs (which when mature are bound by pABP)[317]. Cup promotes deadenylation and protects deadenylated mRNAs from degradation, acting to maintain mRNAs in a repressed state[318].

In addition to proteins normally found in the cytoplasm or in specialised organelles such as p-granules, in a tissue specific manner, proteins involved in chromosome structure and function were also identified. These include SMC2, Top2, Gnf1, glu, Rm62, Nipped-A, Nup153, msps, Dp1, brm, cmet, Dhc64C, row, Mi-2 and Mtor. Also Rif1, woc, piwi, Rad50 and PIAS were identified and are associated with telomeric chromatin. Top2 was identified by 30 or more peptides in both 1-3hr embryos and 12- 15hr embryos. Woc and row were also identified in both conditions by 30 or more peptides (Figure 5–1, Figure 5–2).

In 12-15hr embryos, the boundary element associated factor BEAF-32 was identified as a significant interactor of PIAS by PIAS IP-MS (Figure 5–2) and was not identified in 1-3hr embryos (Figure 5–1).

5.3.2 Identifying PIAS interactors during early *Drosophila* embryonic development

To assess whether PIAS interactors change during early *Drosophila* development, embryos were harvested at more specifically defined timepoints. I aimed to identify how PIAS interactors change over developmental time and possibly identify proteins involved in heterochromatin formation through PIAS-mediated SUMOylation.

Early embryos undergo rapid cycles of cell division, which dilutes the chromatin marks inherited from the parental lineage. In *Drosophila*, chromatin marks are re-introduced as the rapid cycles of cell division slow, nearing the mid-blastula transition and zygotic genome activation[123]. The earliest observed heterochromatin mark is H3K9me2 at nuclear cycle 12[85] with methylation of H3K9 and recruitment of HP1a to chromatin increasing between NC12-14, visible in domains by NC14[123]. PIAS is also observed during these timepoints and is first detected at chromatin in NC10 (Heun lab unpublished).

Embryos were collected at three timepoints early in *Drosophila* development, with the bulk of the embryos at either before (mean at nuclear cycle 9), during (mean at nuclear cycle 12) or after (mean just after nuclear cycles have been completed, in gastrulation) heterochromatin establishment (Figure 5–3A). Representative images of DAPI stained embryos are shown in Figure 5–3B. *Drosophila* embryo protein extracts contain PIAS proteins amenable to immunoprecipitation (Figure 5–3C). Therefore PIAS interaction partners at these different stages in early development could be identified by IP-MS and compared to interactors identified from 12 hour embryos (referred to as timepoint 12), when heterochromatin localisation of PIAS has been lost and chromatin domains have been established.

The $\log_2$ transformed LFQ values of proteins identified to interact with PIAS were plotted as scatter plots (Figure 5–4, Figure 5–5, Figure 5–6) comparing the abundance of proteins identified in PIAS IP-MS from different timepoints. This allowed visualisation of changes in abundance of PIAS interactors that are in common in each timepoint as well as visualization of the number of proteins that

are only identified by PIAS IP-MS to co-immunoprecipitate with PIAS at one timepoint (Figure 5–4, Figure 5–5 and Figure 5–6).

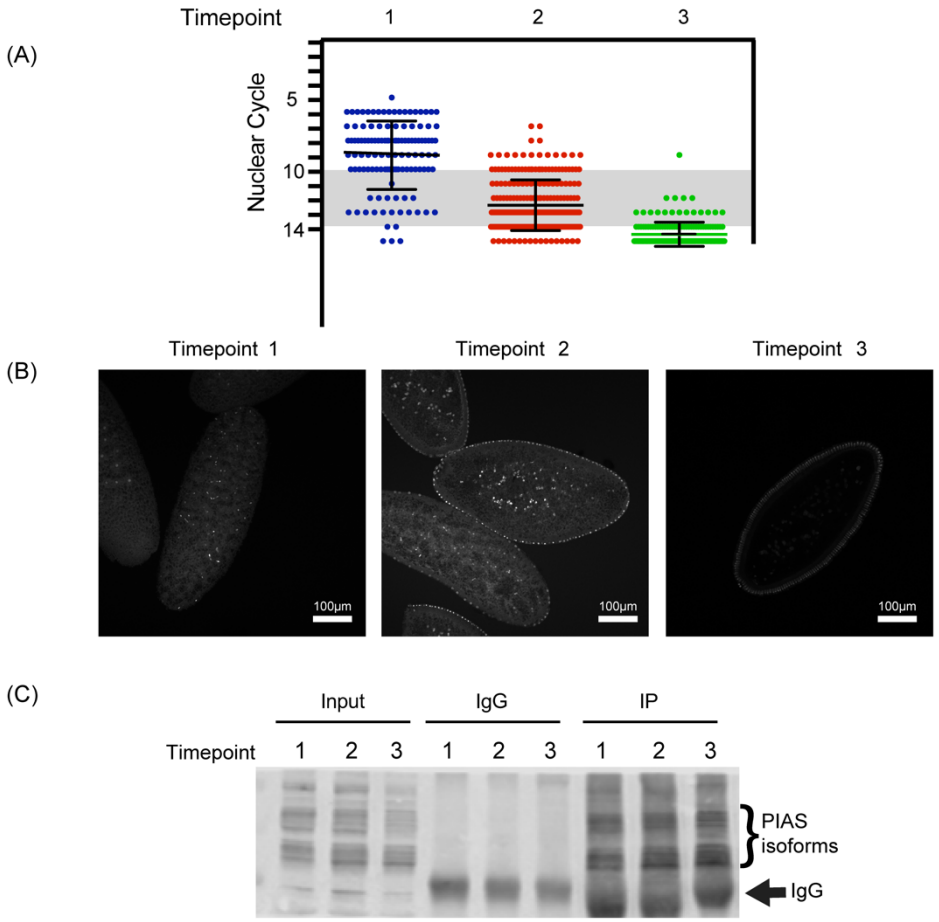


Figure 5–3 Collection of embryos at defined timepoints in development

(A) Embryo nuclear cycles quantified for timepoints 1, 2 or 3, to validate embryo ages of timepoints collected. Embryos were staged by DAPI stain based on nuclear cycles 1-14. n=100/timepoint. Grey shading marks nuclear cycles 10-14 during which heterochromatin is established. Mean +/- standard error **(B)** Representative confocal microscope images of timepoint one, timepoint two and timepoint three DAPI stained embryos. **(C)** Western blot probed with anti-PIAS sheep antibody analysing immunoprecipitations (IP) using affinity purified *Drosophila* anti-PIAS sheep antibody or control anti-IgG sheep antibody coupled to Protein G Dynabeads™ from embryo protein extracts made from ORER *Drosophila* collected timepoints 1-3.

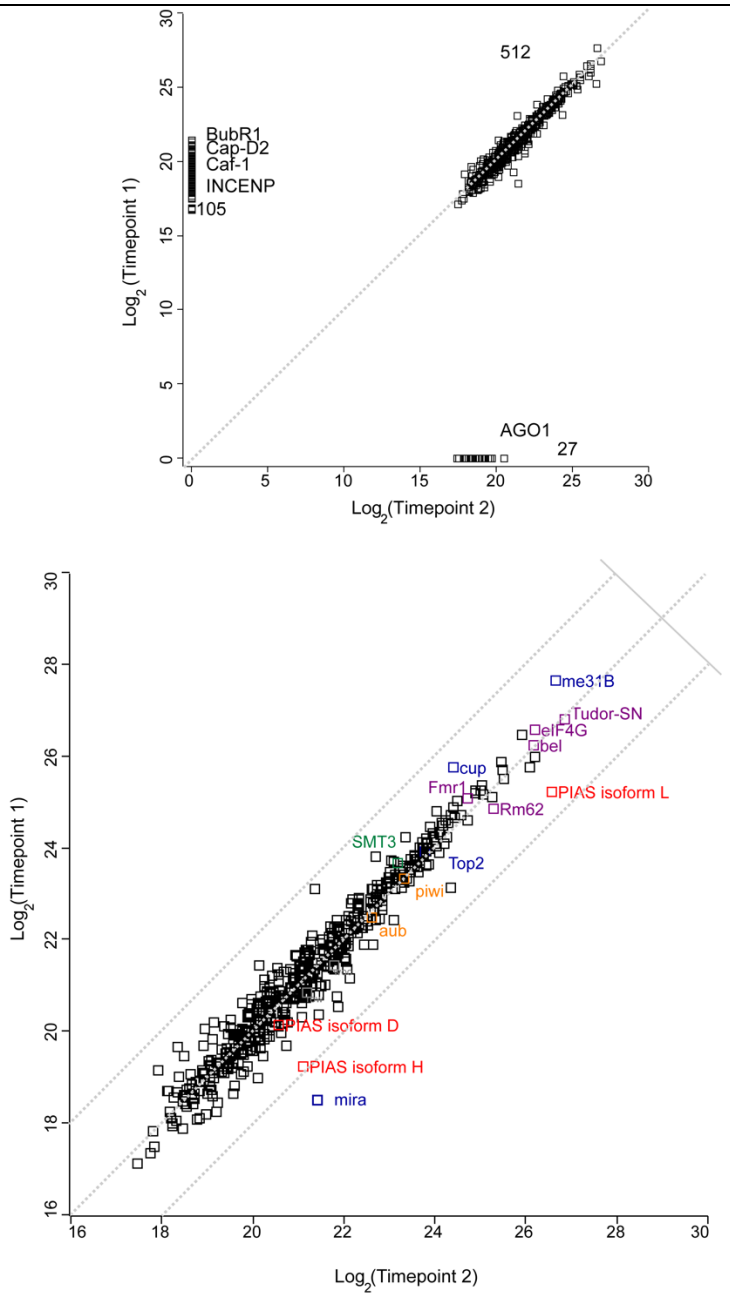


Figure 5–4 Scatter plot comparing PIAS protein interactors in embryos at timepoint 1 and timepoint 2

Mean log_2 transformed LFQ values calculated from three biological replicates, from PIAS IP from timepoint 1 and timepoint 2 embryos in +NEM. Each dot represents one protein. The top plot shows all proteins and the bottom plot shows proteins found only in both conditions. PIAS is in red. In purple are proteins linked to RNAi silencing pathways. In blue are selected proteins which change in interaction specificity during development. In orange are developmental specific RNAi components. SMT3 is in green. Diagonal lines mark two log_2 fold change from zero.

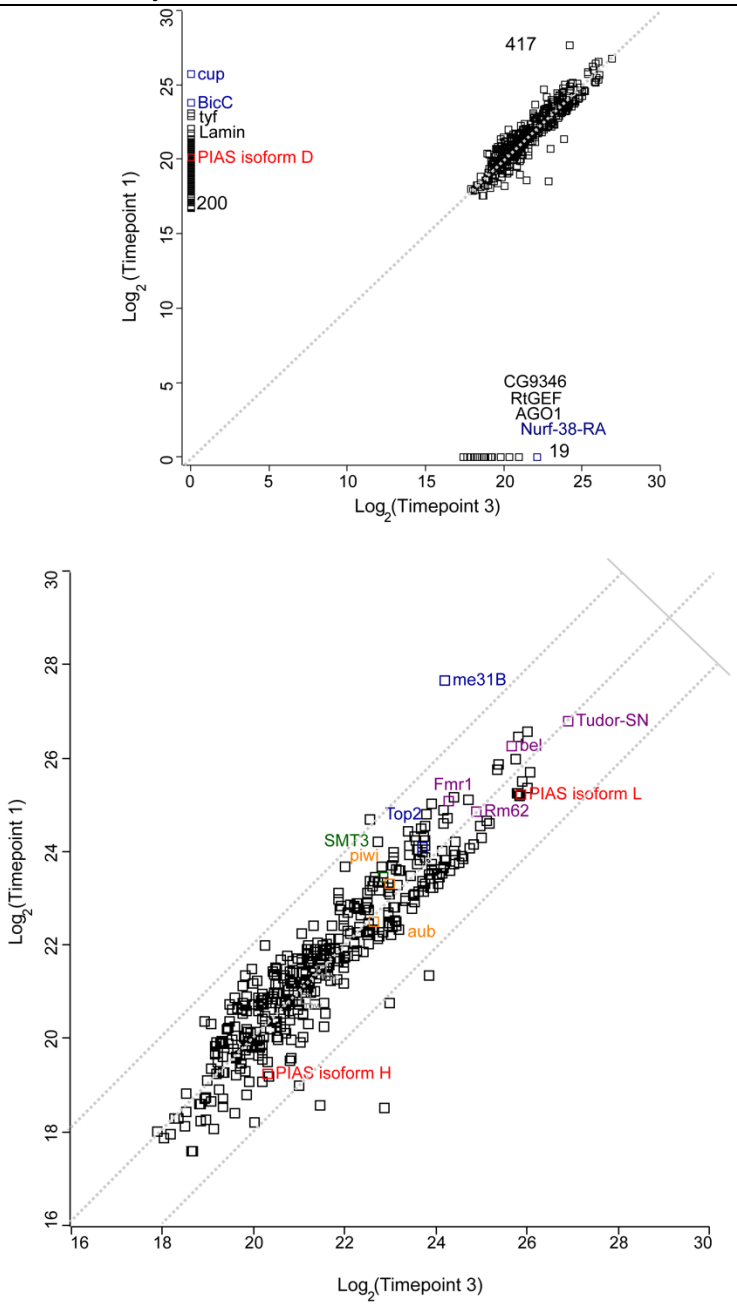


Figure 5–5 Scatter plot comparing PIAS protein interactors in embryos at timepoint 1 and timepoint 3

Mean log_2 transformed LFQ values calculated from three biological replicates, from PIAS IP from timepoint 1 and timepoint 3 embryos in +NEM. Each dot represents one protein. The top plot shows all proteins and the bottom plot shows proteins found only in both conditions. PIAS is in red. In purple are proteins linked to RNAi silencing pathways. In blue are selected proteins which change in interaction specificity during development. In orange are developmental specific RNAi components. SMT3 is in green. Diagonal lines mark two log_2 fold change from zero.

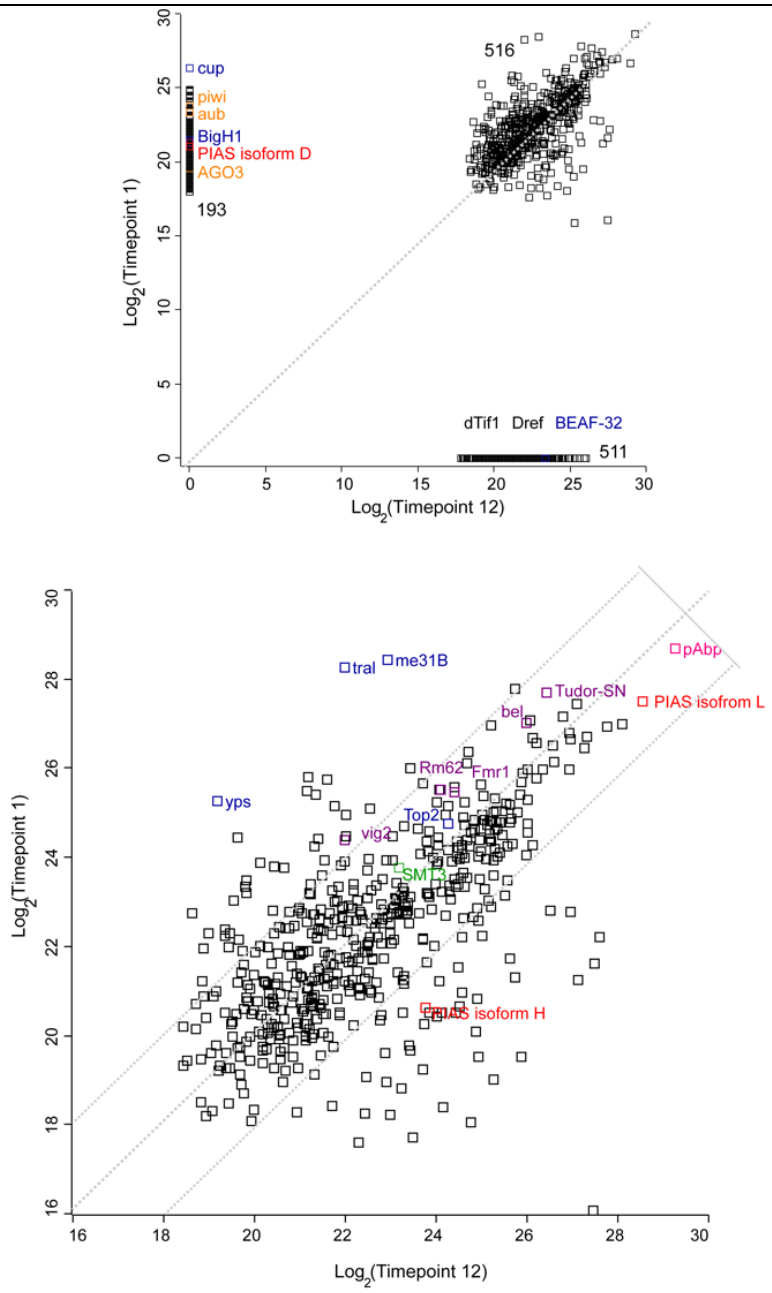


Figure 5–6 Scatter plot comparing PIAS protein interactors in embryos at timepoint 1 and timepoint 12

Mean log_2 transformed LFQ values calculated from three biological replicates, from PIAS IP from timepoint 1 and timepoint 12 embryos in +NEM. Each dot represents one protein. The top plot shows all proteins and the bottom plot shows proteins found only in both conditions. PIAS is in red. In purple are proteins linked to RNAi silencing pathways. In blue are selected proteins that change in interaction specificity during development. In orange are developmental-specific RNAi components. SMT3 is in green. Diagonal lines mark two log_2 fold change from zero.

No differences greater than log_2 two-fold were observed between shared PIAS interaction partners between timepoints 1 and 2 (Figure 5–4) and timepoints 1 and 3, with the exception of me31B which was detected in higher abundance in 1hr compared to 3hr embryos (Figure 5–5).

There are a small number of differences between timepoints 1 and 2 (Figure 5–4), with 512 proteins found at both timepoints and 132 proteins unique to either timepoint. The 105 proteins found to interact with PIAS in timepoint 1 embryos include mitotic checkpoint control protein kinase BUB1 (BubR1), chromatin assembly factor 1, (Caf1) and inner centromere protein (INCENP). Along with 26 other proteins, AGO1 is present in timepoint 2 onwards, and has been shown to have activity at later timepoints in development. AGO1 functions in an RNA induced silencing pathway to silence microRNAs and is important for patterning in the embryo[319].

Between timepoints 1 and 3, the chromatin component Nurf-38 was only found to interact with PIAS in timepoint 3 embryos. Interestingly, cup, bicC and PIAS isoform D are not identified in timepoint 3 (Figure 5–5).

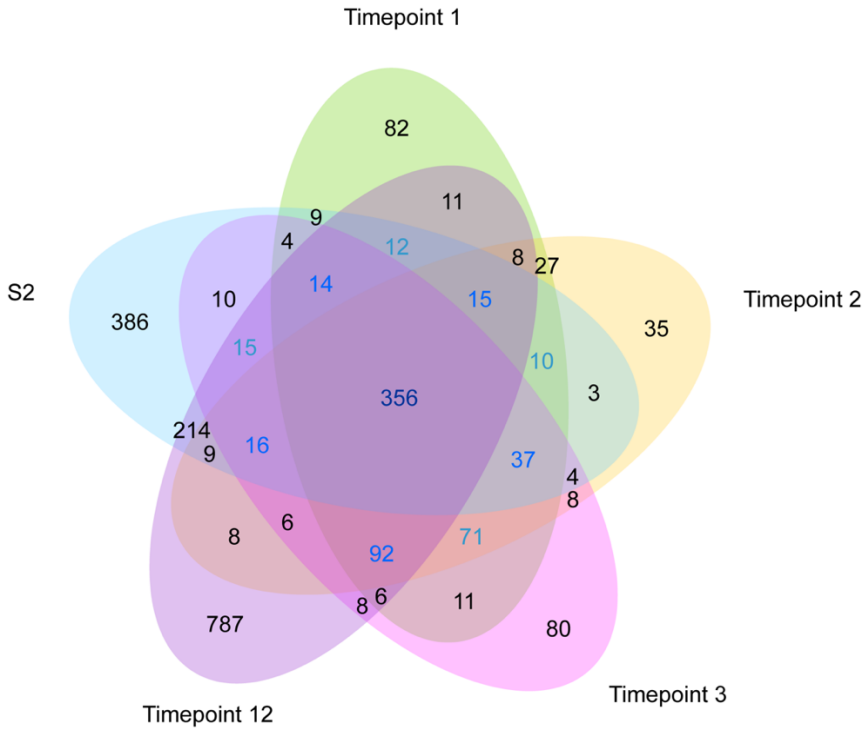
Differences greater than log_2 two-fold were observed between LFQ values of interactors co-immunoprecipitating with PIAS from embryos harvested at timepoints 1 and 12 (Figure 5–6). 516 proteins were present at both timepoints and 704 proteins unique to either timepoint. BEAF- 32 was only identified as a PIAS interactor in timepoint 12 embryos. In Figure 5–6 piwi and aub were present in timepoint 1 not timepoint 12 likely due to their specific developmental role.

5.3.3 Comparison of PIAS interactors identified by IP-MS from *Drosophila* embryos and S2 cells

Proteins identified by two or more peptides in PIAS IP-MS from S2 cells and embryos at timepoints 1,2,3 and 12, in the presence of NEM conditions were compared. The number of overlapping UniProt protein IDs identified are shown in a five-way Venn diagram (Figure 5–7). Negative control samples were included in this analysis, where protein lysate was incubated with IgG-coupled beads. Any protein identified by more than one peptide in the negative control samples was excluded from the overlap as a non-specific interactor of antibody-bead complexes.

A comparison between interactors identified from S2 cells, embryos at 12-15hr and embryos at 1-3hr in NEM conditions found 357 proteins in common between the IP-MS from embryos and S2 cells and 552 interactors in common between S2 cells and 12-15hr embryos (Figure 5–7B) once tissues have been specified[320].

A



B

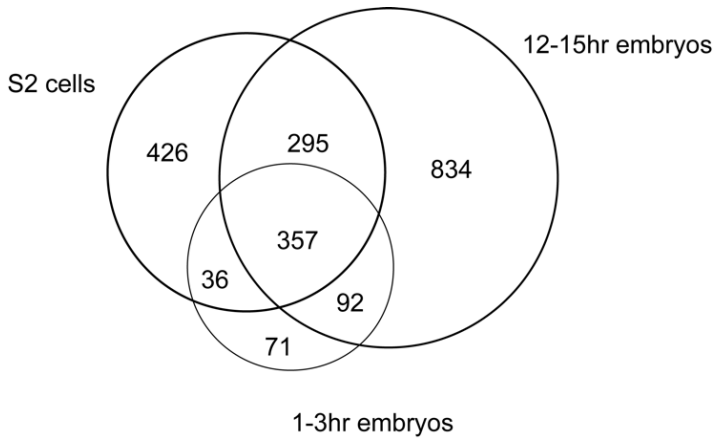


Figure 5–7 Venn diagrams showing the number of proteins identified in PIAS IP-MS
PIAS IP-MS experiments **(A)** from S2 cells and embryos at timepoints 1,2, 3 and 12 **(B)** S2 cells, embryos at 12-15hr and embryos at 1-3hr. All experiments were conducted in the presence of NEM.

5.3.4 Gene ontology (GO) terms analysis of PIAS interactors found in embryos and S2 cells in the presence of NEM

Protein IDs were imported into FlyMine (http://www.flymine.org/) to categorise PIAS interactors based on Gene Ontology (GO) terms, homology searches and pathway enrichment. Of the 357 proteins identified as PIAS interactors in S2 cells, embryos at 1-3hr and embryos at 12-15hr (Figure 5–7B), 331 had GO terms associated with them in FlyMine. The top GO terms associated with PIAS binding partners in S2 cells also found in 1-3hr and 12-15hr embryos are listed in Table 9.

Table 9 Gene ontology (GO) terms associated with proteins identified in all PIAS IP-MS

Gene Ontology analysis, using the FlyMine complete GO biological process annotation dataset, of proteins identified in PIAS IP-MS from S2 cells, embryos at 1-3 and 12-15hr. Bonferroni test correction was applied. Matches are unique proteins within each GO term. In bold are terms of interest.

GO term: Biological Process	p-value	matches
cytoplasmic translation [GO:0002181]	1.07E-73	76
gene expression [GO:0010467]	**3.89E-48**	**234**
cellular nitrogen compound metabolic process [GO:0034641]	3.12E-41	245
chromatin assembly or disassembly [GO:0006333]	**2.33E-38**	**54**
chromosome organization [GO:0051276]	**3.45E-33**	**104**
translation [GO:0006412]	4.75E-31	107
peptide biosynthetic process [GO:0043043]	9.60E-30	108
amide biosynthetic process [GO:0043604]	2.29E-28	108
organelle organization [GO:0006996]	2.33E-27	165
cellular macromolecule biosynthetic process [GO:0034645]	3.57E-27	180
cellular nitrogen compound biosynthetic process [GO:0044271]	1.29E-26	179
macromolecule biosynthetic process [GO:0009059]	1.65E-26	180
peptide metabolic process [GO:0006518]	7.73E-26	108
cellular process [GO:0009987]	**1.10E-24**	**354**
cellular component organization or biogenesis [GO:0071840]	**1.60E-23**	**207**
chromatin organization [GO:0006325]	**1.63E-23**	**75**
cellular amide metabolic process [GO:0043603]	3.22E-23	109
protein-containing complex subunit organization [GO:0043933]	**3.46E-22**	**88**
protein-containing complex assembly [GO:0065003]	**7.43E-22**	**79**
cellular protein-containing complex assembly [GO:0034622]	**1.52E-21**	**75**

cellular component organization [GO:0016043]	2.41E-20	195
DNA-templated transcription, initiation [GO:0006352]	**6.45E-20**	**35**
nucleic acid metabolic process [GO:0090304]	7.03E-20	158
mRNA metabolic process [GO:0016071]	1.07E-19	63
macromolecule metabolic process [GO:0043170]	2.41E-19	264
RNA metabolic process [GO:0016070]	**5.19E-19**	**146**
cellular biosynthetic process [GO:0044249]	1.08E-18	186
cellular component biogenesis [GO:0044085]	1.52E-18	122
RNA processing [GO:0006396]	**2.32E-18**	**75**
organic substance biosynthetic process [GO:1901576]	4.14E-18	187
biosynthetic process [GO:0009058]	1.13E-17	187
DNA conformation change [GO:0071103]	**2.53E-17**	**41**
mRNA processing [GO:0006397]	3.18E-17	54
nucleobase-containing compound metabolic process [GO:0006139]	5.10E-17	164
DNA packaging [GO:0006323]	**1.05E-16**	**38**

In order for a GO terms category to be assigned, the protein ID has to be recognised in the database and have that GO term associated with it. The same list of PIAS interactors (found in all three settings: in S2 cells and embryos at 1-3hr and 12-15hr (Figure 5–7B)), was imported into PANTHER (http://www.pantherdb.org/) for gene ontology (GO) term analysis to categorise the interactors and validate results from FlyMine. Of the 357 proteins, 201 had GO terms associated with them. The top GO terms associated with PIAS interactors are listed in Table 10, ordered by fold enrichment. Fold enrichment is the observed value over the expected value. A value greater than one indicates that the category is overrepresented.

Table 10 Top gene ontology (GO) terms associated with proteins identified in PIAS IP-MS in NEM conditions

Gene Ontology analysis of proteins identified in PIAS IP-MS from S2 cells, embryos at 1-3 and 12-15hr. The analysis is a PANTHER overrepresentation test against the complete GO-Slim biological process annotation dataset with Bonferroni correction applied.

PANTHER GO-Slim Biological Process	Fold Enrichment	p-value
negative regulation of mitotic cell cycle	22.83	3.34E-02
mitotic cell cycle checkpoint	22.83	3.34E-02
ribonucleoprotein complex assembly	10.9	4.72E-03
chromosome organization	9.97	5.55E-08
chromatin organization	8.81	1.17E-03
RNA processing	7.07	1.73E-06
mRNA splicing, via spliceosome	6.85	8.71E-04
RNA splicing, via transesterification reactions	6.79	9.50E-04
translation	6.68	3.93E-08
RNA splicing	6.61	1.23E-03
formation of translation initiation ternary complex	6.58	1.53E-06
translational termination	6.58	1.53E-06
translational elongation	6.58	1.53E-06
cellular component organization or biogenesis	6.19	3.21E-05
cellular component biogenesis	6.19	3.21E-05
ribonucleoprotein complex biogenesis	5.42	7.93E-03
nucleic acid metabolic process	4.14	4.67E-03
organelle organization	3.96	1.32E-05
gene expression	3.48	1.13E-11
macromolecule metabolic process	3.13	1.49E-13
cellular component organization	2.52	6.94E-03
organic substance metabolic process	2.14	1.10E-06
metabolic process	1.88	7.28E-05

Cell cycle processes were identified in FlyMine with a p-value of 0.001659 and 47 independent matches. Proteins involved in the mitotic cell cycle include Bub1-related kinase (A1Z6I7), DNA topoisomerase 2 (P15348), Bub1 kinase isoform A (Q9VMS5), Bub3 isoform A (Q9VAJ2). Interactors of PIAS that are involved in chromatin remodelling include Caf1, H2A, H2B, FACT complex dre4, Iswi. More specifically involved in heterochromatin assembly are Su(var)2-HP2, Dp1, Fmr1 and HP1a. Nuclear pore protein Nup98-96 was identified. Nup98-96 plays a role in chromosome organisation by tethering genes transcribed by RNA Pol II at the nuclear periphery and promotes transcription activation by influencing promoter-enhancer looping. Structural maintenance of chromosomes 2 (SMC2), glu (part of the Condensin complex), and Topoisomerase 2 (Top2) involved in chromosome condensation were also identified. The protein “will die slowly” (wds) that catalyses methylation of histone H3-K4 was also identified.

Classification of GO terms by protein type using PANTHER found that out of 148 molecular function hits, nucleic acid binding was identified on 71 occasions. Hydrolases (proteases, deacetylases, phosphatases) and transferase (acetyltransferase, kinase, methyltransferase) were the next most commonly identified (Figure 5–8).

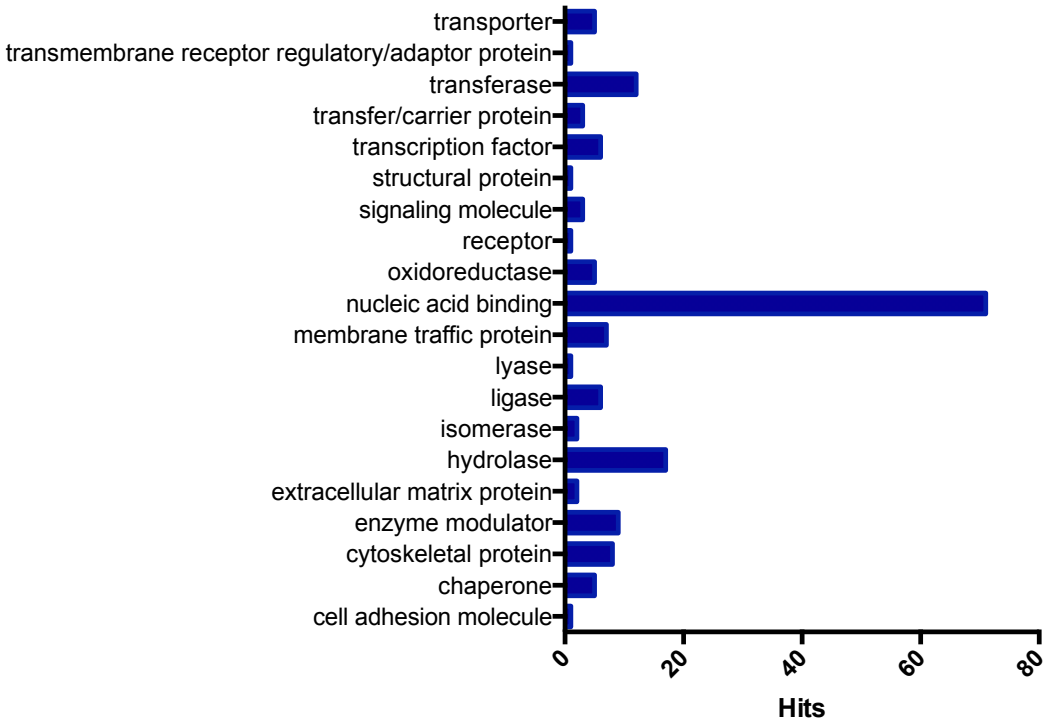


Figure 5–8 PIAS interactors categorised into PANTHER protein class level 1
Proteins identified in PIAS +NEM conditions from S2 cells, embryos at 12-15hr and at 1-3hr were input into The PANTHER Classification System (www.pantherdb.org) for assigning protein class based on GO terms.

Of the nucleic acid binding proteins, 54 proteins were classified as binding to RNA and 14 were classified as DNA binding proteins (Figure 5–9).

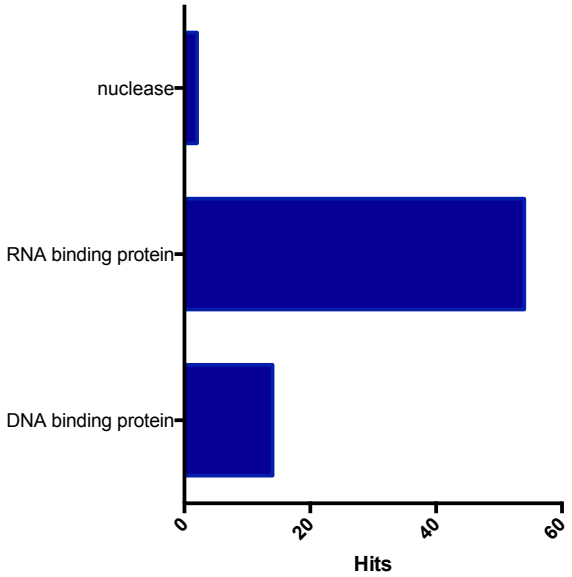


Figure 5–9 PIAS interactors categorised into PANTHER protein class level 2, for nucleic acid binding

The DNA binding proteins included DNA polymerases and chromatin-binding proteins (Figure 5–10).

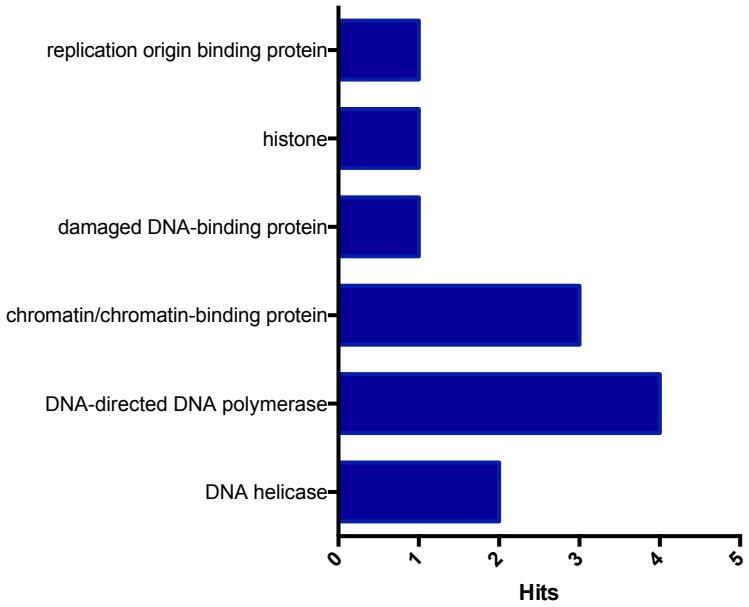


Figure 5–10 PIAS interactors categorised into PANTHER protein class level 3, for DNA binding

Of the RNA binding proteins ribosomal proteins were represented by 20 different proteins (Figure 5–11). Ribosomes are assembled in the nucleus and ribosomal proteins are highly abundant in the nucleus. 14 mRNA processing proteins were identified (Figure 5–11), 10 of which are involved in mRNA splicing and two are involved in mRNA polyadenylation.

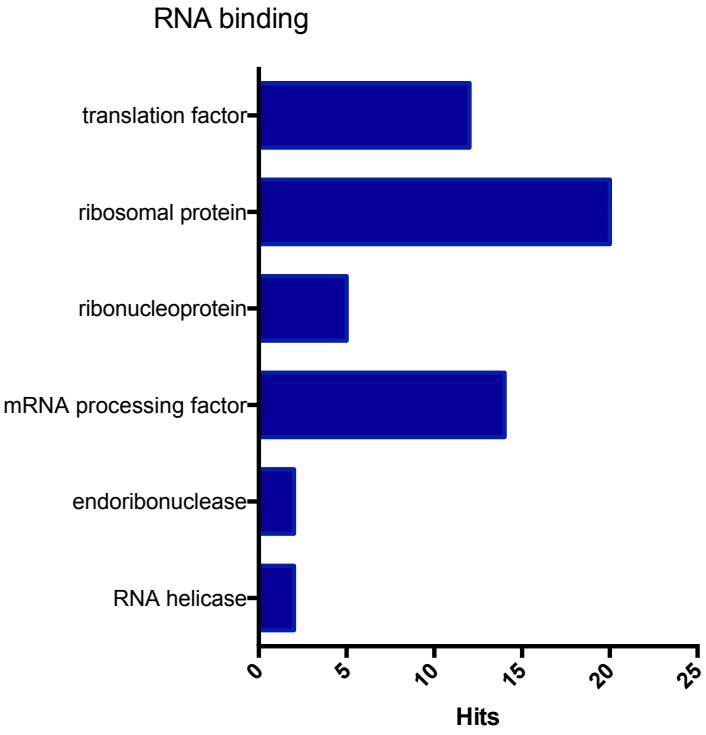


Figure 5–11 PIAS interactors categorised into PANTHER protein class level 3, for RNA binding

5.3.5 PIAS interactors found only in early embryos

Next, to identify proteins that interacted with PIAS specifically in early *Drosophila* embryos (when heterochromatin is being established), I annotated the PIAS interactors found uniquely in timepoint 1 (Table 11) and timepoint 2 (Table 15) embryos, and the interactors found at both of these timepoints (Table 13). My rationale was that if PIAS only localises at heterochromatin in *Drosophila* embryos, and in S2 cells binds euchromatin, something that interacts with PIAS in the early stages of development before heterochromatin establishment could be important for this process. I also queried these protein lists in a GO terms analysis (Table 12 and Table 14).

GO terms enrichment of PIAS interactors from timepoint 1 embryos found terms associated with replication, cell cycle and chromosome segregation (Table 12). For proteins which interact with PIAS in timepoint 1 and 2 embryos, proteins involved in gamete generation and reproductive processes were enriched (Table 14). No enrichment for GO terms was found from timepoint 2 embryos.

Table 11 PIAS interacting proteins unique to timepoint 1 embryos identified by PIAS IP-MS

	Mapped UniProt ID	Gene Name	Protein name	Function
1	A0A0B4K713	fl(2)d	Female lethal d	mRNA methylation, regulation of alternative splicing
2	A0A0B4K730	ced-6	Ced-6	PTB domain-containing adapter protein
3	A0A0B4KFV0	nopo	No poles	DNA polymerase binding, mitotic cell cycle regulation
4	A0A0B4KHU1	Rlip	Ral interacting protein	GTPase activator activity
5	A0A0B4KHY8	CG10254		Ubiquitin-protein transferase activity
6	A0A0B4LG06	gem	Gemini	DNA binding transcription factor activity
7	A0A0B4LGC0	TppII	Tripeptidyl-peptidase II	Serine-type endopeptidase activity
8	A0A0H4Y1G0	Nelf-A	FI24705p1	Negative elongation factor
9	A1Z6H4	vlc	RE52822p	signaling
10	B7Z0A9	sowah	Sosondowah	
11	B7Z138	CG34408		GTPase activator
12	C0MIY3	CG2867		
13	E8NH59	WRNexo	RE48478p	3’-5’ exonuclease activity
14	H1UUH1	eco	FI18257p1	Sister chromatid cohesion and meiotic chromosome segregation
15	L0CRU3	mael	Maelstrom	Regulation of gene silencing by miRNA
16	M9NDA6	Lasp	Lasp	
17	M9PCF1	Duox	Dual oxidase	Response to oxidative stress
18	M9PFF9	tral	Trailer hitch	Cytoskeleton organisation

19	M9W9W1	Chd3	FI21135p1	ATP binding chromatin remodeller
20	O18475	mus308	DNA polymerase theta	Promotes non-homologous end-joining DNA
21	O76912	CG2701	EG:95B7.9	(high expression in eye disks)
22	P20439-3	CycB-RB	G2/mitotic-specific cyclin-B	Cell cycle regulation
23	P25158-2	osk	Maternal effect protein oskar	Organises the germ plasm
24	P25171	Rcc1	Regulator of chromosome condensation	Promotes the exchange of Ran-bound GDP by GTP. Involved in the regulation of onset of chromosome condensation in the S phase
25	P54622	mtSSB	Single-stranded DNA-binding protein, mitochondria	This protein binds preferentially and cooperatively to pyrimidine rich ss-DNA. Required for mitochondrial DNA replication.
26	P54733-2	CycE	G1/S-specific cyclin-E	Cell cycle regulation
27	Q05913	TfIIFalpha	General transcription factor IIF subunit 1	TFIIF is a general transcription initiation factor that binds to RNA polymerase II and helps to recruit it to the initiation complex in collaboration with TFIIB.
28	Q0KHS7	Lsd-2	Lipid storage droplet-2	
29	Q24154	RpL29	60S ribosomal protein L29	chromatin binding[321], structural constituent of ribosome
30	Q24558	tos	Exonuclease 1	5'->3' double-stranded DNA exonuclease which may also contain a cryptic 3'->5' double-stranded DNA exonuclease activity
31	Q29QV6	N/A	IP14020p	DNA repair
32	**Q32KD2**	**egg**	**Eggless**	**Histone-lysine N-methyltransferase, trimethylates histone 3 lysine 9**
33	Q59DT6	Rb97D	Ribonuclear protein at 97D	Germ cell development and mRNA binding
34	Q5U118	AMPKalpha	Non-specific serine/threonine protein kinase	
35	Q5UEC2	Taf12L	TAF-like protein	Transcription initiation from RNA polymerase II promoter
36	Q7JUM5	MCPH1	LD43341p	Mitotic cell cycle
37	Q7JXA2	PRAS40	LOBE	Negative regulation of protein kinase activity
38	Q7KJQ4	IKKepsilon	Protein kinase DIK2	protein kinase

39	Q86NY8	Tomosyn	LD44990p	GTPase activator activity, cell polarity establishment or maintenance
40	Q8IPM0	srl	Spargel	mRNA binding, transcription coregulator activity, energy homeostasis
41	Q8IR24	sun	Protein stunted	Activates the G-protein coupled receptor mth *in vitro*, leading to increased intracellular calcium ion levels.
42	Q8IRG9	Iml1	GATOR complex protein Iml1	TORC1 pathway regulation, to mediate metabolic homeostasis, female gametogenesis
43	Q8T389	EndoB	Endophilin B	Membrane organisation
44	Q960Y4	drosha	LD29995p	siRNA production for RNA interference
45	Q9I7N6	mus201	DNA repair endonuclease	Double-strand break repair and nucleotide-excision repair
46	Q9I7U4-3	sls	Titin	Muscle protein
47	Q9U3V9	xmas-2	Xmas-2	GERMINAL-CENTER ASSOCIATED NUCLEAR PROTEIN Required for spermatogenesis, oogenesis and embryogenesis, mRNA export from nucleus
48	Q9U8Q0	px	Plexus	Imaginal disc-derived wing vein morphogenesis
49	Q9V9V7	pasha	LD23072p	Double-stranded RNA binding, gene silencing by miRNA
50	Q9VBA3	BcDNA:GH07466	AT30002p	
51	Q9VD66	fit	Female-specific independent of transformer	
52	Q9VDR1	MED25	Mediator of RNA polymerase II transcription subunit 25	Component of the Mediator complex, a coactivator involved in the regulated transcription of nearly all RNA polymerase II-dependent genes
53	Q9VEH1	Mps1	LD04521p	Protein kinase, chromosome segregation
54	Q9VFP2-3	rdx;	Protein roadkill	Involved in segment polarity
55	Q9VI72	Sas-4	Spindle assembly abnormal 4	centromere protein J, centrosome separation
56	Q9VIS5	CG10631	Uncharacterised protein	Negative regulation of transcription by RNA polymerase II
57	Q9VK08	CG44085	Uncharacterised protein	MAP-kinase scaffold activity

58	Q9VKH9	cana	CENP-ana, isoform A	CENTROMERE-ASSOCIATED PROTEIN E metaphase/anaphase transition of mitotic cell cycle
59	Q9VL20	ova	RE35195p	core promoter sequence-specific DNA binding
60	Q9VLR5	Ssb-c31a	RNA polymerase II transcriptional coactivator	General coactivator that functions cooperatively with TAFs and mediates functional interactions between upstream activators and the general transcriptional machinery
61	Q9VM45	Nuf2	Nuf2	Chromosome segregation
62	Q9VMX6	Marcal1	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1	ATP-dependent annealing helicase that catalyzes the rewinding of the stably unwound DNA
63	Q9VNE4	asl	Asterless	Asymmetric cell division
64	Q9VNR6	l(3)04053	Lethal (3) 04053	
65	Q9VP12	SMC5	Structural maintenance of chromosomes 5	double-strand break repair via homologous recombination protein sumoylation
66	Q9VP94	CG10588	Uncharacterised protein	
67	Q9VPY0	IntS14	Integrator complex 14	Component of the Integrator complex, a complex involved in the transcription of small nuclear RNAs (snRNA)
68	Q9VS48	mus312	Structure-specific endonuclease subunit SLX4	resolving diverse forms of deleterious DNA structures
69	Q9VSP0	CG6511	IP13783p	
70	Q9VUI0	gnu	CG5272	Protein kinase activator activity, mitotic cell cycle regulation
71	Q9VX63	CG8915-RA	FI18001p1	ATP-DEPENDENT RNA HELICASE YTHDC2-RELATED N6-methyladenosine-containing RNA binding
72	Q9VXA0	Nprl2	GATOR complex protein NPRL2	TORC1 signaling pathway
73	Q9VYR0	Lsm12	LSM12	Component of the Atx2-tyf activator complex
74	Q9VZ04	GCS1	Glucosidase 1	Protein N-linked glycosylation

75	Q9W028	yellow-g2	IP19120p	Cuticle pigmentation, melanin biosynthetic process
76	Q9W4W7	EG:100G10.6	RE68603p	POLYCOMB PROTEIN SCM
77	Q9W5Y4	Marf1	MIP18541p	Regulator of gene expression
78	Q9XZ03	mbf1	Multiprotein Bridging Factor 1	DNA binding, central nervous system development
79	T2FFI3	mei-9	FI21265p1	DNA repair
80	X2J6U8	ssp3	Short spindle 3	cytoskeleton
81	X2JKF2	CG9213	Uncharacterised protein	
82	X2JKM9	kat80	Katanin p80 WD40 repeat-containing subunit B1	Microtubule binding

Table 12: GO terms analysis of the 82 PIAS interacting proteins unique to timepoint 1 embryos identified by PIAS IP-MS

GO term: Biological process	p value	matches
nuclear chromosome segregation [GO:0098813]	0.001056	10
DNA replication [GO:0006260]	0.001726	8
chromosome segregation [GO:0007059]	0.002525	10
cell cycle [GO:0007049]	0.013219	17
regulation of DNA replication [GO:0006275]	0.018052	5

Table 13 PIAS interacting proteins shared in timepoint 1 and timepoint 2 embryos identified by PIAS IP-MS

	Mapped UniProtID	Gene Name	Protein Name	Function
1	A0A0B4KGX7	gwl	Greatwall, isoform B	chromosome condensation, negative regulation of phosphoprotein phosphatase activity
2	A0A0B4KGY9	msi	Musashi, isoform H	translation repressor activity, mRNA regulatory element binding, cell fate determination
3	A0APA5	CG3509	CG3509 protein	DNA binding, nucleosome assembly
4	A1A708	CG4951	Uncharacterized protein	
5	A1Z6E6	d4	D4, isoform C	histone acetyltransferase activity, transcription regulation
6	A1ZAP7	CG6967	Uncharacterized protein	posttranscriptional gene silencing by RNA
7	G7H828	CG9247-RA	RE72821p1	3'-5' exonuclease activity
8	M9NDR2	BicD	Bicaudal D, isoform B	cytoskeletal adaptor activity
9	M9PHL6	bif	Bifocal, isoform E	axon guidance, female meiosis chromosome segregation
10	O46112	toc	Toucan protein	syncytial blastoderm mitotic cell cycle
11	O96828	Psi	EG:EG0003.2	mRNA processing
12	Q4V527	CG15415	IP13307p	
13	Q7KSP6	Sbf	SET domain binding factor, isoform B	histone acetyltransferase and methyltransferase activity, transcription regulation
14	Q86RA0	Mbs	Myosin phosphatase DMBS-L	phosphatase regulator activity
15	Q8I9J8	scaf6	SR-related CTD associated factor 6	RNA processing
16	Q8IGK3	Su(var)2-10	RE73180p, isoform D	
17	Q8INW9	fon	Fondue, isoform A	metamorphosis

18	Q8MR83	muskelin	AT11715p	regulation of cell-matrix adhesion
19	Q9NH11	RecQ4	RECQ4	ATP-dependent 3'-5' DNA helicase activity, telomere maintenance, DNA repair and replication
20	Q9VBI4	Exo84	Exocyst 84, isoform A	establishment or maintenance of cell polarity
21	Q9VDE3	slmb	LD08669p	phosphoprotein binding
22	Q9VQM4	gkt	Probable tyrosyl-DNA phosphodiesterase	DNA repair enzyme
23	Q9VV79	spd-2	BcDNA.LD24702	phosphohydrolase activity
24	Q9VW68	Gabat	Gamma-aminobutyric acid transaminase, isoform A	metabolic process
25	Q9W3C4	Hexo2	Beta-hexosaminidase	metabolic process
26	X2JC02	CG45186	Supervillin	
27	X2JEK3	wisp	Wispy, isoform B	polynucleotide adenylyltransferase activity

Table 14 GO terms analysis of the 27 PIAS interacting proteins shared in timepoint 1 and timepoint 2 embryos

GO term: Biological process	p-value	matches
multi-organism process [GO:0051704]	0.001913	15
reproduction [GO:0000003]	0.008982	14
reproductive process [GO:0022414]	0.015306	13
cellular process involved in reproduction in multicellular organism [GO:0022412]	0.017972	11
multicellular organism reproduction [GO:0032504]	0.022163	13
multi-organism reproductive process [GO:0044703]	0.026886	12
gamete generation [GO:0007276]	0.04792	11

Table 15 PIAS interacting proteins unique to timepoint 2 embryos identified by PIAS IP-MS

	Mapped UniProtID	Gene Name	Protein Name	Function
1	A0A0B4LFQ7	CG9646-RA	Uncharacterized protein	
2	A1Z6G3	Fis1	Mitochondrial fission 1	Involved in the fragmentation of the mitochondrial network and its perinuclear clustering
3	B6IDY8	Nprl3	FI07656p	negative regulation of TOR signaling
4	B7Z026	PDZ-GEF	Dizzy isoform CII	Ras GTPase binding, germ-line stem cell population maintenance, imaginal disc-derived male genitalia development, imaginal disc-derived wing morphogenesis, regulation of compound eye photoreceptor development
5	D1Z363	Cp36	RT04026p	
6	E1JHK3	dl	Dorsal	DNA-binding transcription factor activity
7	M9PBB5	vas	Vasa	intracellular mRNA and protein localization
8	M9PHK2	Tao	Tao	MAP kinase kinase kinase activity
9	**O17468-3**	**Hira**	**HIRA homolog**	**DNA replication-independent nucleosome assembly, transcription corepressor activity, sperm chromatin decondensation, regulation of chromatin silencing**
10	P42283	lola	Longitudinals lacking protein	Putative transcription factor required for axon growth and guidance in the central and peripheral nervous systems
11	Q24169	Orc5	Origin recognition complex subunit 5	Component of the origin recognition complex (ORC) that binds origins of replication
12	**Q59E36-4**	**CoRest**	**REST corepressor**	**represses transcription of neuron-specific genes in non-neuronal cells**
13	Q6NR30	BEST:LD13441	RE38958p	
14	Q7JRH2	ReepA	Receptor expression-enhancing protein	
15	Q7JRM5	CG11127	GH12965p	
16	Q7JW66	CG12129	LD21545p	

17	Q7K1D9	CG15512	GH14951p	
18	Q7YU52	E23	RE53253p	regulation of circadian rhythm
19	Q8IGH0	CG5484-RB		
20	Q8MQR6	CG7029 P	RH23428p	
21	Q8SYF6	Cyp4s3	RE63964p	
22	Q8SYX1	CG32576	RE29988p	
23	Q8SZD0	Grip71		gamma-tubulin binding, mitotic cell cycle
24	Q8T921	CG11710	AT04454p	transcription coactivator activity
25	Q95V48	CCY	Coiled-coil Y protein	
26	Q9V3E9	spag	FI17138p1	negative regulation of motor neuron apoptotic process
27	Q9V8K2	Sec6	Exocyst complex component 3	SNARE binding
28	Q9VE79	CG14309	Uncharacterized protein	
29	Q9VG32	Men	Malic enzyme	determination of adult lifespan
30	Q9VHN4	CG8043	GH14121p	
31	Q9VNC2	HDAC3	Histone deacetylase	histone H3 and H4 deacetylation
32	Q9VPX4	Tfb4	FI01003p	RNA polymerase II CTD heptapeptide repeat kinase activity
33	Q9VSV2	CG4476	Transporter	neurotransmitter transport
34	Q9VZ69	Imp	IGF-II mRNA-binding protein	neuron remodeling
35	Q9W158	CG4612	LD36772p	long-term memory, positive regulation of translation

5.3.6 XNP and ADD1 (ATRX) as interactors of PIAS in embryos

As the Heun lab is interested in the role of PIAS mediated SUMOylated of chromatin associated factors, I focused on XNP and ADD1. XNP had previously been observed to co-localise with DAPI dense regions of pre-blastoderm nuclei in *Drosophila* embryos[144], coinciding with the timings of heterochromatin formation and it was identified as a SUMO target in both a study in the Heun lab by a previous student[304] and independently by Nie et al., 2009[171]. ADD1 was of interest because it interacts with XNP[296] and could possibly aid chromatin recognition of XNP if acting in a similar manner to its mammalian ATRX homolog.

I had shown that ADD1 interacts with PIAS in S2 cells and that the PIAS-XNP interaction only occurred in the presence of NEM, suggesting that that interaction depends on SUMOylation. Whether an interaction between XNP and ADD1 is observed in *Drosophila*, using the same tools as used in S2 cells, might clarify whether ADD1-XNP is maintained *in vivo*.

XNP is moderately expressed in the embryo, whereas ADD1 has very low expression levels, (Appendix Figure 9–6)[139]. Endogenous XNP was not detectable above background levels in *Drosophila* embryos when stained with the anti-XNP rabbit antibody (generated in this study- see 6.3.6) (data not shown). However, the anti-ADD1 rabbit antibody (generated in this study- see 6.3.6) recognised ADD1 epitopes in *Drosophila* embryos despite ADD1 having lower expression levels than XNP[139]. This suggests that the anti-ADD1 antibody is more sensitive than the anti-XNP antibody, which is not surprising given that XNP immunization schedule was terminated early, likely before the time when the antibodies have reached their optimum recognition of targets after a certain number of boosts.

ADD1 co-localises with DAPI dense regions of pre-blastoderm nuclei in the *Drosophila* embryo (Figure 5–12). Interestingly, ADD1 is found at discrete sites (Figure 5–12), that likely correspond to telomere ends as ADD1, XNP and PIAS bind telomeric DNA sequences of chromosome 2L[322] and ADD1 and HP1a co-localise at the tips of the chromosome arms in wild-type polytene chromosomes[295].

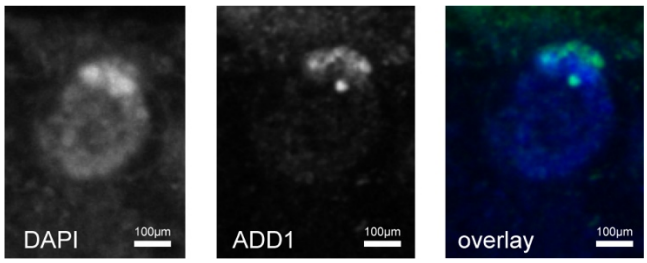


Figure 5–12 ADD1 localises at heterochromatin in nuclear cycles 12-14 in *Drosophila* embryos.

Representative confocal image of a pre-blastoderm cell in a *Drosophila* embryo at nuclear cycle 13. Embryos were stained with anti-ADD1 antisera and DAPI.

Localisation of PIAS and ADD1 was analysed in *Drosophila* embryos. PIAS and ADD1 co-localise with DAPI-dense regions at discrete sites (Figure 5–13).

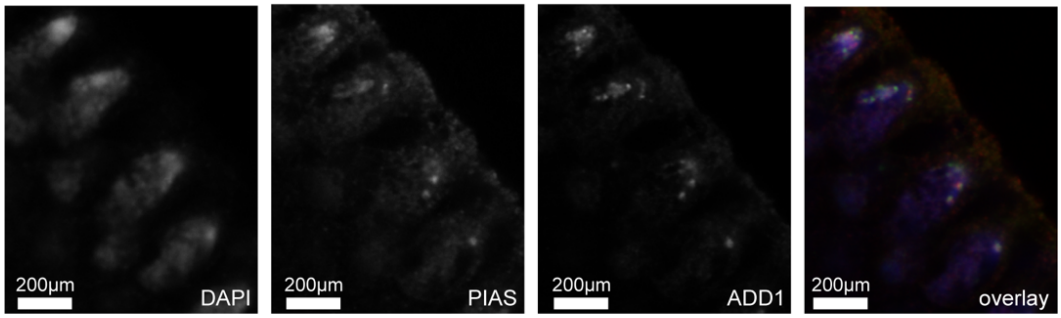


Figure 5–13 PIAS and ADD1 colocalise at heterochromatin in nuclear cycle 13-14 *Drosophila* embryos.

Representative confocal image of pre-blastoderm cells in *Drosophila* embryos at nuclear cycle 14 (NC14). Embryos were stained with affinity purified anti-PIAS sheep antibody (red), anti-ADD1 antisera (green) and DAPI.

ADD1 and XNP were not identified as interactors of PIAS in *Drosophila* embryos by IP-MS. ADD1 was detected by one peptide in PIAS IP-MS from 12-15hr embryos, however two or more peptides of a protein need to be detected to identify a protein in mass spectrometry studies. XNP was not identified. However, PIAS IP followed by western blot probing with anti-ADD1 and anti-XNP antibodies showed that PIAS, ADD1 and XNP co-immunoprecipitate from 1-3hr embryos and 12-15hr embryos (Figure 5–14).

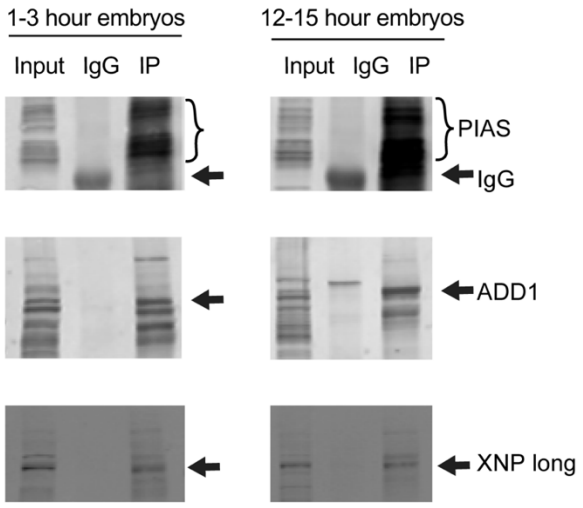


Figure 5–14 ADD1 and XNP co-immunoprecipitate with PIAS in *Drosophila* embryos. Protein extracts were made from *Drosophila* embryos. The input (IN), IP with affinity purified anti-IgG sheep antibody (IgG) and IP with anti-PIAS sheep antibody (IP) were analysed on 8% SDS-PAGE followed by western blot probed with *Drosophila* anti-PIAS sheep antibody, anti-ADD1 antibody or anti-ATRX rat antibody.

5.4 Summary

To identify common functions of PIAS interactors, proteins identified in PIAS IP-MS from S2 cells (section 4.3.9) and from embryos collected at different timepoints (Figure 5–3) were compared. This revealed a common set of PIAS interactors identified in all conditions (Figure 5–7), as-well as a number of proteins that co-immunoprecipitated with PIAS either specifically in either embryos or S2 cells. GO terms analysis by PANTHER overrepresentation test and identification of associated GO terms of PIAS interactors allowed inference of the biological processes that PIAS interactors are involved in.

5.4.1 Common functions of PIAS interactors

PIAS interactors are enriched with nucleic acid binding proteins (Figure 5–8). The nucleic acid binding proteins represented interaction with DNA binding proteins and RNA binding proteins, in particular, chromatin proteins, DNA polymerases, ribosomal proteins and mRNA processing proteins (Figure 5–9, Figure 5–10, Figure 5–11). Overall, this suggests that PIAS mediated SUMOylation is involved in different stages of gene expression regulation. However, the exact role of PIAS mediated SUMOylation is as yet unknown. Experiments removing the E3 ligase activity of PIAS would need to be performed to assign the function of PIAS interaction in aiding SUMOylation of these proteins.

5.4.2 PIAS interactors only present in embryos

In PIAS IP-MS from timepoint 1 embryos, Eggless was identified (Table 11). Eggless is of interest in the context of heterochromatin establishment in *Drosophila* embryos (see section 2.1.1). The mechanism of Eggless recruitment to chromatin is currently unknown and is hypothesised to be recruited to chromatin through interaction with small RNAs. It is plausible that, like HP1α in mouse tissue culture cells (discussed in section 2.3.13), SUMOylation could be involved in Eggless-RNA guided recruitment to establish heterochromatin.

If SUMOylation is involved in establishment of processes, proteins which require SUMO may only display an interaction with PIAS transiently at a specific timepoint. For example, processes such as sister chromatid cohesion[323,324], kinetochore

assembly[325,326] and centromere identification[327] require SUMO mediated processes. Immunofluorescence staining in early *Drosophila* embryos has identified co-localization of PIAS and the centromere identifier protein (CID) uniquely in early embryos (Heun lab unpublished). Identification of INCENP as an interactor of PIAS supports a role of PIAS mediated SUMOylation in the establishment of centromere protein localisation at centromeric chromatin regions.

5.4.3 ADD1 and XNP co-immunoprecipitate with PIAS in *Drosophila* embryos

During nuclear cycles (NC) 12-14, ADD1 is apically localised in pre-blastoderm nuclei at DAPI dense regions (Figure 5–12), in a similar pattern to SMT3[328], PIAS[144], XNP[144] and H3.3[144], coinciding with the timepoint for *de novo* establishment of heterochromatin. Although ADD1 and XNP co-immunoprecipitate with PIAS in embryos (Figure 5–14), neither protein was identified by two or more peptides in PIAS IP-MS. There are several potential reasons for this. Firstly, ADD1 is expressed at low levels[142] and might be outcompeted by other interactors for binding to PIAS. Secondly, many ADD1 bands were detected on western blot in 1-3hr embryos (Figure 5–14), which could mean alterative degradation products or post-translational modifications, which results in ADD1 peptides not being identified in IP-MS experiments.

5.4.4 PIAS co-immunopreciptates with Piwi and aub in early embryos and PIAS co-immunopreciptates RNAi pathway components

I was interested in confirming the interaction between PIAS and Piwi as the Aravin and Toth labs had identified PIAS as an interactor of Piwi[316] and have proposed that PIAS interacts with Eggless[126], providing a link between the Piwi/piRNA silencing pathway and proteins required for deposition of H3K9me3 at chromatin and recruitment of HP1a to form heterochromatin domains to silence regions of the genome which contain transposable elements[126]. Piwi and aub were found to interact with PIAS at timepoint 1 but not timepoint 12 (Figure 5–6), likely due to their specific developmental role. SUMOylation may be essential for localisation of proteins to carry out Piwi-piRNA mediated transcriptional repression and to establish heterochromatin. Similarly, SUMOylation by PIAS may play a similar role in related RNAi pathways as RNAi components were also identified to interact with PIAS (Figure 5–1 and Figure 5–2).

PIAS co-immunoprecipitates with bel, AGO2, Tudor-SN and FMR1 (Figure 5–1, Figure 5–2). This supports a role for PIAS (possibly through PIAS mediated SUMOylation) in influencing interactions with components of the RNA-induced silencing complex (RISC) to aid their function in an RNAi mediated pathway that is active throughout *Drosophila* development in tissues and in cell lines. SUMOylation of AGO2 has been investigated[243]. AGO2 can be SUMOylated by PIAS proteins, with SUMO protecting AGO2 from degradation and mediating AGO2 association with Dicer[243]. Therefore it is possible that PIAS mediated SUMOylation may regulate RNAi pathway activity through aiding RISC complex formation.

5.4.5 PIAS has non-chromatin associated interactions

PIAS is also likely to be involved in regulation of post-transcriptional gene silencing (PTGS) in the early embryo. The protein with the greatest difference in LFQ value between PIAS IP and control was polyadenylate-binding protein (pAbp) (Figure 5–1 and Figure 5–2). This binds poly(A) tails of mature mRNAs, and signals that the transcript has properly matured, marking it for translation and protects mRNAs from degradation. pAbp interacts with the translation initiation factor eIF4G to stimulate translation[329].

Proteins found in the nuage/p-granule/germ plasm may be SUMOylated by PIAS, which may facilitate SUMO mediated storage or release. For example, to maintain repressive SUMO states or release components when required, as in, a mechanism for FMRP SUMOylation[330]. SUMO mediated mechanisms in the *Drosophila* embryo may be important for the correct timings of protein expression and protein localisation in developmentally regulatory pathways, balancing expression and degradation of mRNA transcripts. Enrichment of cup, eIF4E1 and e1F4G in PIAS IP-MS (Figure 5–1, Figure 5–2) is interesting as cup is a SUMO substrate[171] involved in maternal mRNA regulation[331,332]. By binding to mRNAs, the protein cup can prevent interaction between eIF4E1 and e1F4G[332]. The role of SUMOylated cup remains to be investigated.

PIAS is a maternally deposited protein and is highly abundant in *Drosophila* embryos, with both nuclei-associated and cytoplasmic localisation (Figure 2–4). pAbp, me31B, eIF4E1, tral and cup were identified in PIAS IP-MS from embryos(Figure 5–1). Together these form a ribonucleoprotein complex in *Drosophila* embryos during nuclear cycles 1-11 that bind to maternal mRNAs to downregulate transcript expression during the maternal to zygotic transition. pAbp forms a complex with eIF4G[333], tyf and Atx2 which interacts with Not1 to degrade mRNA transcripts in circadian pacemaker neurones[334]. Atx2 interaction with tyf and Lsm12 activates PTGS, whereas Atx2 interaction with me31B represses PTGS to regulate translation of genes involved in circadian rhythms[335]. eIF4G is SUMOylated in *S. pombe*[336] and in *S. cerevisiae*[337], suggesting a level of SUMO mediated regulation that could be mediated by PIAS in *Drosophila* embryos.

5.4.6 PIAS interacts with insulator proteins after nuclear cycles are complete

BEAF- 32 was only identified as a PIAS interactor in 12-15hr and timepoint 12 embryos. This is of notable interest because BEAF-32 was shown to co-immunoprecipitate with PIAS in S2 cells (Figure 4–18B), which are derived from late stage (20-24hr) embryos. This could suggest a role for PIAS mediated SUMOylation of chromatin insulator proteins. This interactions is only found at later stages in *Drosophila* development, once chromatin states have been defined and is not observed in early stage embryos undergoing nuclear cycle divisions.

5.4.7 Evaluation of the approach

5.4.7.1 Protein extract and IP-MS from timed embryos

Embryos at specific developmental stages were collected. This enabled identification of PIAS interaction partners that otherwise would not have been represented in a mixed or a later pool of embryos (Figure 5–4, Figure 5–5, Figure 5–6). All proteins that are part of PIAS-associated complexes were immunoprecipitated and detected by mass spectrometry. This resulted in a large number of proteins identified, especially in timepoint 12 embryos (compared to timepoints 1,2 and 3) where there are different tissues specified. It is likely that the requirements for PIAS expression and localisation differs between these timepoints. Live imaging of GFP-PIAS in the developing embryo identified PIAS in the

cytoplasm and only showed only partial co-localisation with DAPI stained chromosomes in interphase and early prophase (Figure 2-4). During prometaphase PIAS formed non-chromatin associated discrete structures in the cytoplasm (Figure 2-4), which dissipated after metaphase, which could coincide with the shut-off of translation. However, PIAS is maternally deposited, which results in large amounts of PIAS in *Drosophila* embryos[138,139] , therefore these PIAS-associated complexes could reflect a stored pool of the protein. For *Drosophila* PIAS protein levels assessed through development, determined by quantitative mass spectrometry, see data collated in Casas-Vila *et al.,* 2017[142] (Appendix Figure 9–6). Improved sensitivity in detecting low levels of proteins or use of alternative methods may enable identification of interactors from more specific timepoints, earlier in development or from less starting material. To identify developmental specific roles of PIAS, within different cellular compartments in early embryos would be useful and interesting. The approach I think most suitable for this would be the expression of tagged-PIAS in a specific tissue to allow isolation of PIAS-associated complexes from that tissue. For example, this is important to separate PIAS-associated protein complexes in the oocyte from PIAS-associated protein complexes in the male gamete. Identifying PIAS interactors within a cellular compartment may reduce the detection of proteins that interact with large pools of maternally deposited protein that is merely being storage for controlled release when needed at later points in development.

Due to localisation of PIAS at heterochromatin in embryos[144] and at euchromatin in Kc167 cells[57], I hypothesised that PIAS would have differential functions and therefore change its interactome. However, the PIAS interactome might largely reflect a change in the proteome and not all interactions are necessarily functionally important. For example, if proteins are present but stored until a later timepoint in development, then they might appear to interact with other stored proteins.

5.4.7.2 Identifying unique PIAS interactors in early embryos

Comparison of proteins identified in PIAS IP-MS from *Drosophila* embryos taken at early timepoints in development allowed separation of PIAS interactors that only occur at specific developmental times. After subtracting common PIAS interacting proteins from the list of proteins identified in early embryos, proteins that specifically

interacted with PIAS in *Drosophila* embryos were involved in gamete generation, reproductive processes, replication, cell cycle and chromosome segregation. I would expect proteins involved in these processes to be subject to post-translational modification by SUMOylation to regulate their functions.

5.4.7.3 Comparison of PIAS interactors in S2 cells and embryos

One drawback of GO analysis is that process annotation is dependent on the database used for the search. Proteins or genes in pathways that are more studied have more GO terms assigned to them. This causes an effect similar to confirmation bias where pathways already established are more likely to be noted than those not established, and an increase in false detection as more comparisons (individual GO terms) are being run for some molecular pathways than others.

5.4.7.4 Immunofluorescence in *Drosophila* embryos

The *Drosophila* anti-PIAS sheep antibody can detect PIAS in early embryos, showing enrichment at heterochromatin in NC14 (Figure 5–13). ADD1 also localises to this region (Figure 5–13). Further analysis of the timings and order of appearance have been explored by Matilde Fabe in the Heun lab during the writing of this thesis. XNP bleed 2 antibodies in serum were at too low levels to be used for detection above background and would require concentration and purification before use in immunofluorescence.

5.4.7.5 ADD1 and XNP as interactors of PIAS *Drosophila* embryos

Use of proteomics to determine protein levels in *Drosophila* embryos, as in the study conducted by Casas-Vila et al., 2017[142], identified low expression of ADD1, moderate expression of XNP and high expression of PIAS. It is possible that ADD1 was not detected as an interactor of PIAS in embryos, not only due to its low expression, but due to limitations in the experimental technique. This could be because mass spectrometry studies rely on the detection of peptides that map uniquely to one protein, and are therefore assigned to that protein. The ADD1-XNP-PIAS interaction in *Drosophila* embryos should be explored further, as a conclusion cannot be made about the hypothesis that ADD1 and XNP interaction with PIAS might be required for the establishment of heterochromatin.

Chapter 6 Production and purification of recombinant XNP and ADD1 and validation of antibodies generated against them

6.1 Objectives

SUMOylation failure of XNP and ADD1 was of interest to the Heun lab as a potential mechanism of PIAS' phenotype at chromatin. Therefore, I aimed to purify XNP, ADD1, SMT3 and PIAS proteins to confirm a direct interaction between these proteins and SUMOylation of XNP *in vitro*. In this chapter I describe the strategy that I used to produce and purify recombinant XNP and ADD1. Due to a lack of information on the key interaction residues and post-translational modification sites, we decided to purify full length XNP and ADD1. Neither *Drosophila* protein had been purified before. Unfortunately, these proteins did not form stable species *in vitro* under the conditions tested but the constructs and methods provide a useful starting point for future purification method development.

In addition, there are no commercial antibodies available for ADD1 or XNP. A *Drosophila* rat ATRX antibody (from Andreas Thomae) can detect denatured forms of XNP on western blot but does not recognise native protein for immunofluorescence and immunoprecipitation. Antibodies against ADD1 were generated by Lopez Falcon *et al.,* 2014[296], which recognise a 150kDa band on WB but a 150kDa band is not detected with HA-tagged ADD1, with purified protein or with endogenous protein, leading to a question over the antibodies specificity. Therefore, I produced polyclonal antibodies for ADD1 and XNP from the purified protein aggregates.

6.2 Aims

- To produce and purify recombinant full length ADD1 and XNP
- To identify SUMOylation sites on ADD1 and XNP
- To confirm and characterise an interaction between XNP and PIAS *in vitro*
- To generate and validate antibodies against ADD1 and XNP

6.3 Results

6.3.1 Expression and purification trials for full length XNP and ADD1 protein

Expression constructs for ADD1 and XNP were generated by a ligation independent cloning method, by designing primers for PCR amplification of the target sequence for insertion into insect expression vectors (provided by the Cook laboratory). Sequences were confirmed by Sanger sequencing using primers spanning every couple of hundred base pairs of DNA (Appendix Figure 8–9, Figure 8–10).

Test expression assays were carried out for GST-ADD1-RA and His-XNP-RA protein expression in Sf9 cells. Western blots probed with anti-GST antibodies or a Ni^{2+}-HRP probe detected expression of the GST or the HIS tag respectively (Figure 6–1). GST-ADD1 was detected at ~200kDa, which is higher than its predicted molecular weight. To confirm that the expressed bands were the correct proteins, the pull down samples were run on SDS-PAGE, the gel was Coomassie stained, the bands were excised and the proteins were digested with trypsin. MALDI-TOF analysis of tryptic peptides confirmed the proteins expressed were *Drosophila* ADD1 and XNP by >30% peptide coverage.

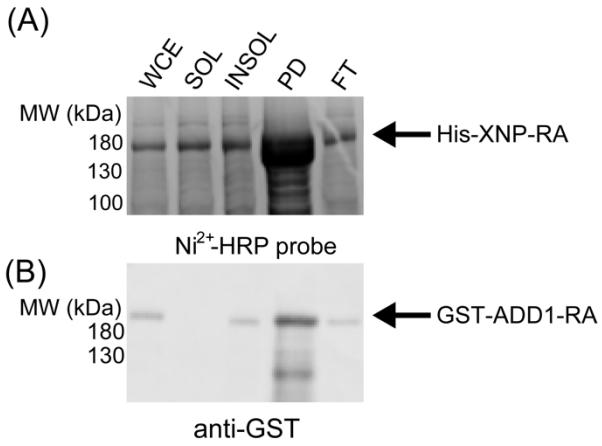


Figure 6–1 Test expression assays for GST-ADD1-RA and His-XNP-RA protein expression in Sf9 cells

(legend continued over page)

(A) Western blot probed with a Ni^{2+}-HRP probe showing the result of a test expression from Sf9 cells infected with the pFastBac-HIS-XNP-RA bacmid. Whole cell extract (WCE), soluble (SOL) and insoluble (INSOL) samples were generated from a 50mL culture of Sf9 cells after 3 days-post-infection. The SOL fraction was subjected to pull down (PD) with Ni^{2+}-NTA resin which recognises the HIS tag on HIS-XNP-RA, resulting in HIS-XNP-RA enrichment in the PD lane at the expected molecular weight of 151kDa. **(B)** Western blot probed with an anti-GST rabbit antibody showing the result of a test expression from Sf9 cells infected with the pFL-GST-ADD1-RA bacmid. Whole cell extract (WCE), soluble (SOL) and insoluble (INSOL) samples were generated from a 50mL culture of Sf9 cells after 3 days-post-infection. The SOL fraction was subjected to pull down (PD) with GSH resin which recognises the GST tag on GST-ADD1- RA, resulting in GST-ADD1-RA enrichment in the PD lane.

After test expressions, several rounds of protein purification were trialled to identify conditions for purification of properly folded ADD1 and XNP proteins.

6.3.2 Purification of HIS-XNP-RA

HIS-XNP-RA was purified from Sf9 cell lysate by batch incubation with Ni^{2+}-NTA affinity resin. Washes removed non-specific proteins resulting in an isolated HIS-XNP-RA protein (Figure 6–2A) and HIS-XNP-RA eluted in 300mM imidazole (Figure 6–2A). The E1 fraction was diluted and analysed on SDS-PAGE (Figure 6–2A). Samples from batch purification were also run on SDS-PAGE and transferred to nitrocellulose membrane for western blot probed with the rat anti-ATRX antibody (Figure 6–2B), confirming the identity of the protein as XNP. Despite running as a single band on SDS-PAGE, XNP apparently did not form single species *in vitro*. HIS-XNP-RA was injected on a Superdex 200 size exclusion column to: separate XNP; to buffer exchange to reduce imidazole concentrations; and to confirm the size of XNP species. XNP eluted close to the void volume of the Superdex 200 column (Figure 6–2C). A size exclusion chromatography column with a different resin composition, Superpose 6 (which has a greater resolving range for higher molecular weight proteins) was trialled to separate XNP. XNP eluted over a broad range (from 10-20mL) (Figure 6–2D), characteristic of high molecular weight complexes, partially folded proteins or aggregates. XNP fractions were concentrated and snap frozen as separate fractions; however, XNP still formed aggregates in solution. This suggests that for XNP to be usable in *in vitro* assays, fresh batches must be prepared.

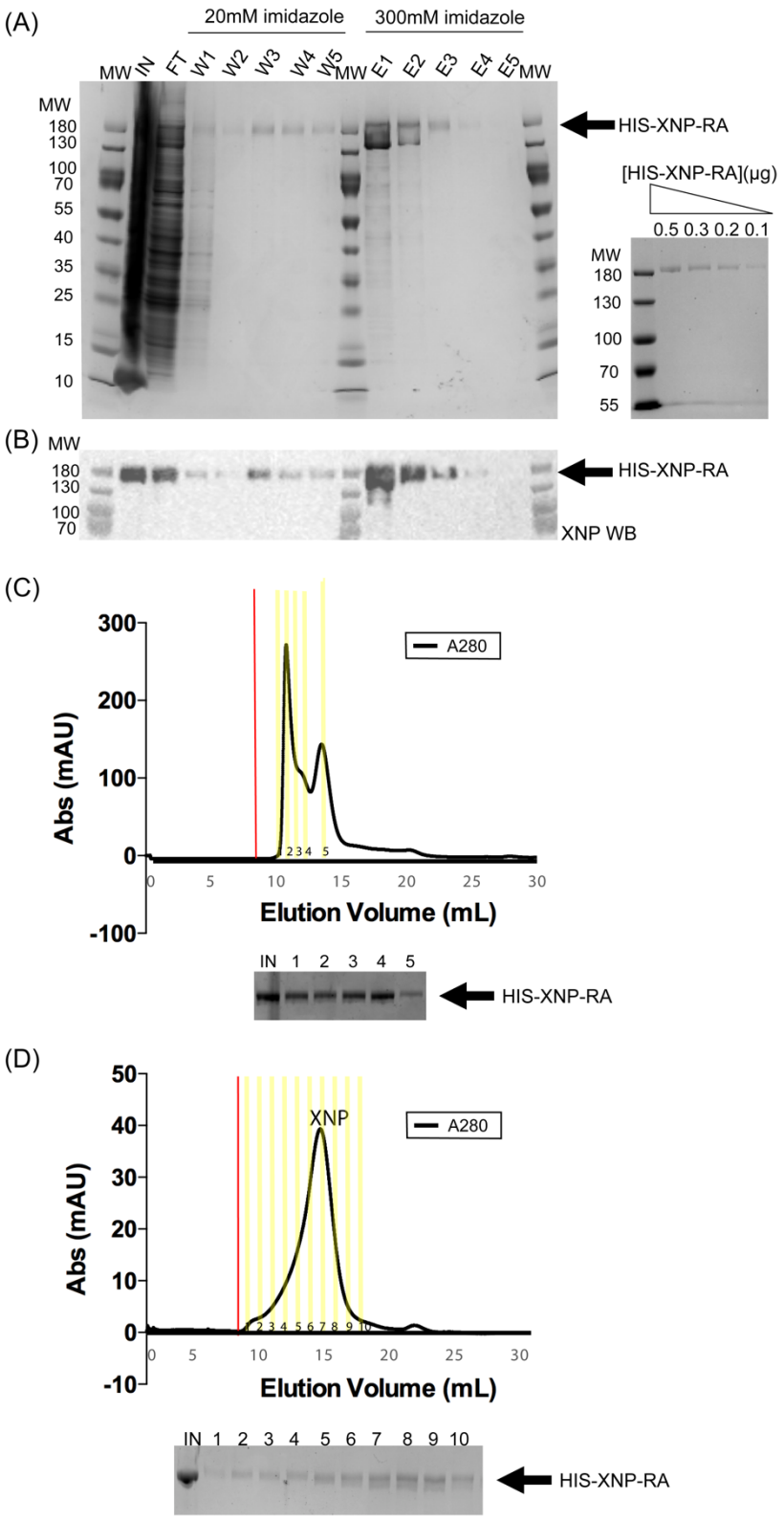

(A)
20mM imidazole
300mM imidazole
MW IN FT W1 W2 W3 W4 W5 MW E1 E2 E3 E4 E5 MW
MW
180
130
100
70
55
40
35
25
15
10
HIS-XNP-RA
[HIS-XNP-RA](μg)
0.5 0.3 0.2 0.1
MW
180
130
100
70
55
(B) MW
180
130
100
70
HIS-XNP-RA
XNP WB
(C)
A280
Abs (mAU)
300
200
100
0
-100
0 5 10 15 20 25 30
Elution Volume (mL)
2 3 4 5
IN 1 2 3 4 5
HIS-XNP-RA
(D)
XNP
A280
Abs (mAU)
50
40
30
20
10
0
-10
0 5 10 15 20 25 30
Elution Volume (mL)
1 2 3 4 5 6 7 8 9 10
IN 1 2 3 4 5 6 7 8 9 10
HIS-XNP-RA

Figure 6–2 Summary of HIS-XNP-RA purification from Sf9 cells

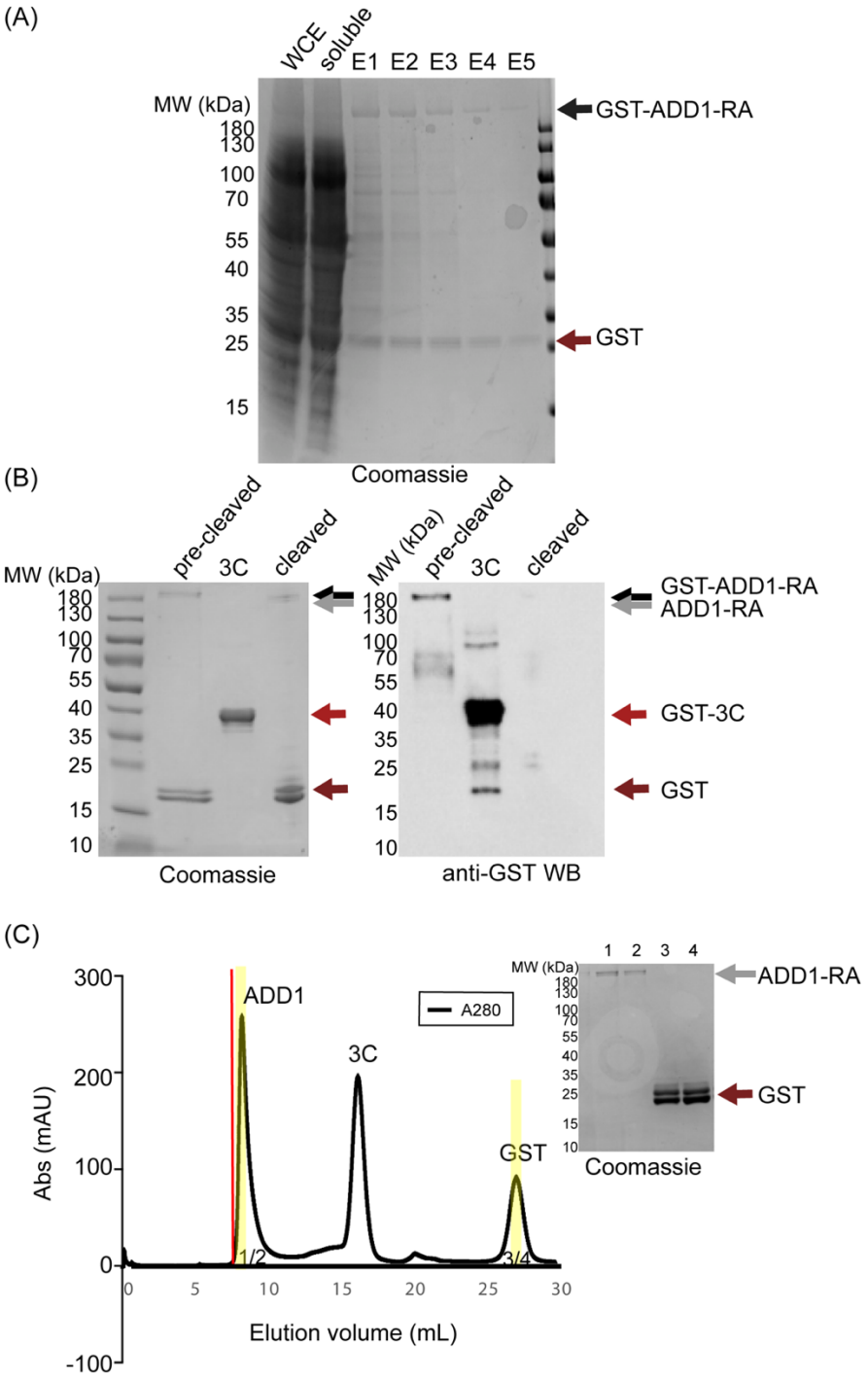
(A) Coomassie stained SDS-PAGE gel showing the input (soluble fraction applied to the resin), FT (from batch binding to Ni-NTA resin), washes in low imidazole (20mM) and elution from Ni^{2+}-NTA resin with high imidazole (300mM) washes. Coomassie stained SDS-PAGE gel loaded with different concentrations of the elution fraction E1 (from A) showing HIS-XNP-RA running above the 180 kDa molecular weight marker on denaturing SDS-PAGE **(B)** The same samples as in A analysed by SDS-PAGE followed by western blot probed with anti-ATRX (XNP) rat antibody to confirm expression of XNP-HIS-RA. The arrow marks the expected size of HIS-XNP-RA **(C)** A280 elution trace of HIS-XNP-RA from a Superdex 200 size exclusion chromatography column, below is a Coomassie stained SDS-PAGE gel taking protein sample from the fractions highlighted in yellow. Elution fractions 1-5 were pooled and frozen. The red line marks the void volume of the column. **(D)** A280 elution trace of HIS-XNP-RA from a Superose 6 size exclusion chromatography column from an analytical run, below is a Coomassie stained SDS-PAGE gel taking protein sample from fractions 1-10 highlighted in yellow. The red line marks the void volume of the column. In all panels the marker is PageRuler™ Prestained Protein Ladder, 10 to 180 kDa (Thermo Fisher Scientific).

6.3.3 Purification of ADD1-RA

GST-ADD1-RA protein was expressed (Figure 6–3A) and purified from Sf9 cell lysate by batch incubation with GSH affinity resin. GST-ADD1-RA bound to the resin and was eluted by competition with excess reduced glutathione (Figure 6–3B). The GST tag was removed by cleavage with GST-3C protease (Figure 6–3C). Once isolated from its tag, ADD1-RA was injected on a Superdex 200 size exclusion column to separate the ADD1-RA, GST-3C protease and free GST by size exclusion chromatography (Figure 6–3D). The elution profile contains three peaks representative of these three species and Coomassie stained SDS-PAGE of selected elution fractions shows the separation of ADD1-RA from GST (Figure 6–3D). Elution fractions containing ADD1-RA were pooled and concentrated. ADD1-RA eluted close to the void volume of the column, indicating that the protein did not interact with the column resin and eluted as an aggregate.

Figure 6–3 Summary of ADD1-RA purification from Sf9 cells

(A) Coomassie stained SDS-PAGE gel showing the soluble fraction applied to the GSH resin and elutions (E1-4) from batch binding to GSH resin. Free GST is at 26kDa and GST-ADD1-RA is above the 180kDa marker. **(B)** Cleavage of the GST tag from GST-ADD1-RA by addition of GST-3C, Left-hand-side: Coomassie stained SDS-PAGE gel showing pre-cleaved GST-ADD1-RA and free GST at 26kDa in lane 1, GST-3C in lane 2, and cleaved GST-ADD1-RA and free GST in lane 3. Right-hand-side: western blot probed with anti-GST antibodies. **(C)** A_{280} elution trace from a Superdex 200 size exclusion chromatography column separating ADD1-RA from 3C and GST which are shown as three peaks. Coomassie stained SDS-PAGE gel shows separation of ADD1-RA (lanes 1 and 2) from free GST (lanes 3 and 4) taking protein sample from the labelled fractions. The red line marks the void volume of the column. In all panels the marker is PageRuler™ Prestained Protein Ladder, 10 to 180 kDa (Thermo Fisher Scientific).


(A)
WCE
soluble
E1 E2 E3 E4 E5
MW (kDa)
180
130
100
70
55
40
35
25
15
GST-ADD1-RA
GST
Coomassie
(B)
pre-cleaved
3C
cleaved
MW (kDa)
GST-ADD1-RA
ADD1-RA
GST-3C
GST
Coomassie
anti-GST WB
(C)
ADD1
3C
GST
A280
Abs (mAU)
Elution volume (mL)
1/2
3/4
1 2 3 4
ADD1-RA
GST
Coomassie

6.3.4 XNP and ADD1 purified protein characterisation

Upon thawing, XNP was observed on Coomassie stained SDS-PAGE (Figure 6–4) as higher molecular mass species. This suggested that the protein might be forming aggregates that migrate through the gel at a slower rate than expected. The HIS-XNP-RA purification protocol was adapted to use high salt and TCEP as a reducing agent throughout the purification procedure, resulting in the same outcome.

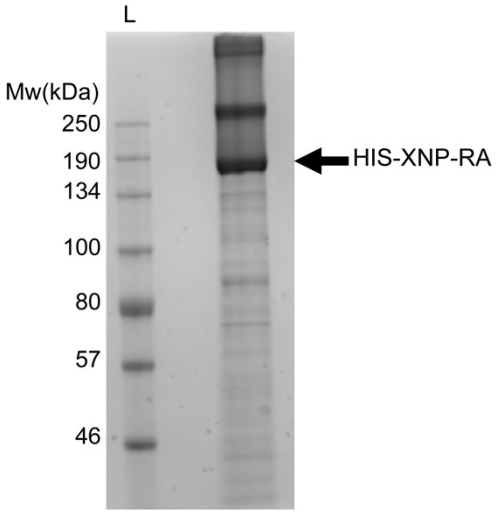


Figure 6–4 SDS resistant XNP-RA bands

Coomassie stained 6% SDS-PAGE gel showing recombinant full-length HIS-XNP-RA protein, eluted from pool 2 (the peak) of XNP fractions after size exclusion chromatography, after freeze thawing. The marker (L) is 250 kDa Color-coded Prestained Protein Marker, Broad Range (CST).

Intriguingly, ADD1-RA runs at a higher molecular weight than expected based on its predicted molecular weight on SDS-PAGE under reducing conditions (Figure 6–5). This may be due to a variety of factors such as the proteins isoelectric point (pI), the amount of SDS bound to the protein or the addition of post-translational modifications. ADD1-RA runs between 190kDA and 200kDa on semi-native, tris-glycine and bis-tris gels. Denaturation in urea buffer and longer boiling in SDS buffer did not alter the protein mobility (data not shown).

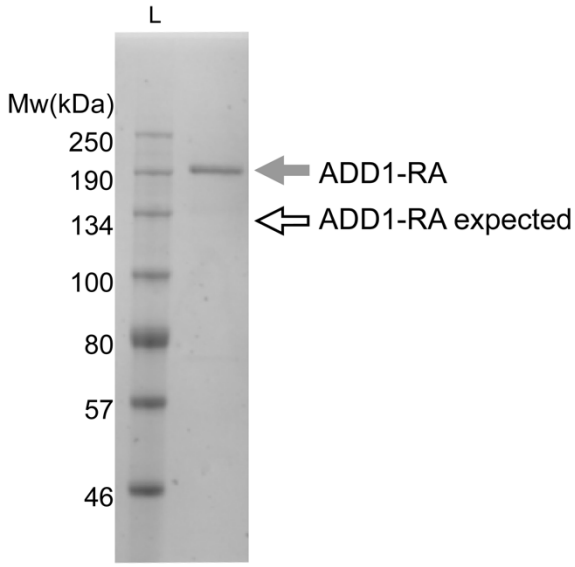


Figure 6–5 ADD1-RA runs at a higher than expected molecular weight
Coomassie stained 6% SDS-PAGE gel showing recombinant full-length ADD1-RA protein (grey arrow) which has an expected size of 131 kDa (open arrow). The marker (L) is 250 kDa Color-coded Prestained Protein Marker, Broad Range (CST).

Dynamic light scattering (DLS) gives a measure of particle size in suspension by detecting photons from scattered laser light projected a through a sample in suspension at short consecutive intervals. The speed of particles undergoing Brownian motion is used to calculate particle size using the Stokes-Einstein equation. DLS was used to check for protein aggregation in the HIS-XNP-RA and ADD1-RA purified protein samples in solution, as larger particles diffuse more slowly in solution and can give an estimate of sample uniformity and size. HIS-XNP-RA had a size distribution between 10-100 d.nm in NaCl concentrations between 0.1-0.5M (Figure 6–6). HIS-XNP-RA was polydisperse in solution. This may also be responsible for a lower (0.5) correlation coefficient value at 0.1µs for HIS-XNP-RA, which is indicative of a low signal to noise ratio of scattered light detected in the HIS-XNP-RA sample over background. Ionic strength, surface structure and shape can all influence these readings.

The large particles in the ADD1-RA sample could be broken up by the increase in salt concentration from 200mM to 500mM (Figure 6–6). In 500mM NaCl, ADD1-RA existed in solution as 10% monodisperse and 90% polydisperse; indicating that the sample was not homogeneous. From the monodisperse peak, the mean particle size (z-average hydrodynamic diameter) of 320kDa was derived from the intensity weighted mean diameter derived from the cumulant analysis. However, it is not possible to determine the oligomeric state through this assay as DLS bases its size estimations on a number of assumptions, the main assumption being that the protein in solution is globular.

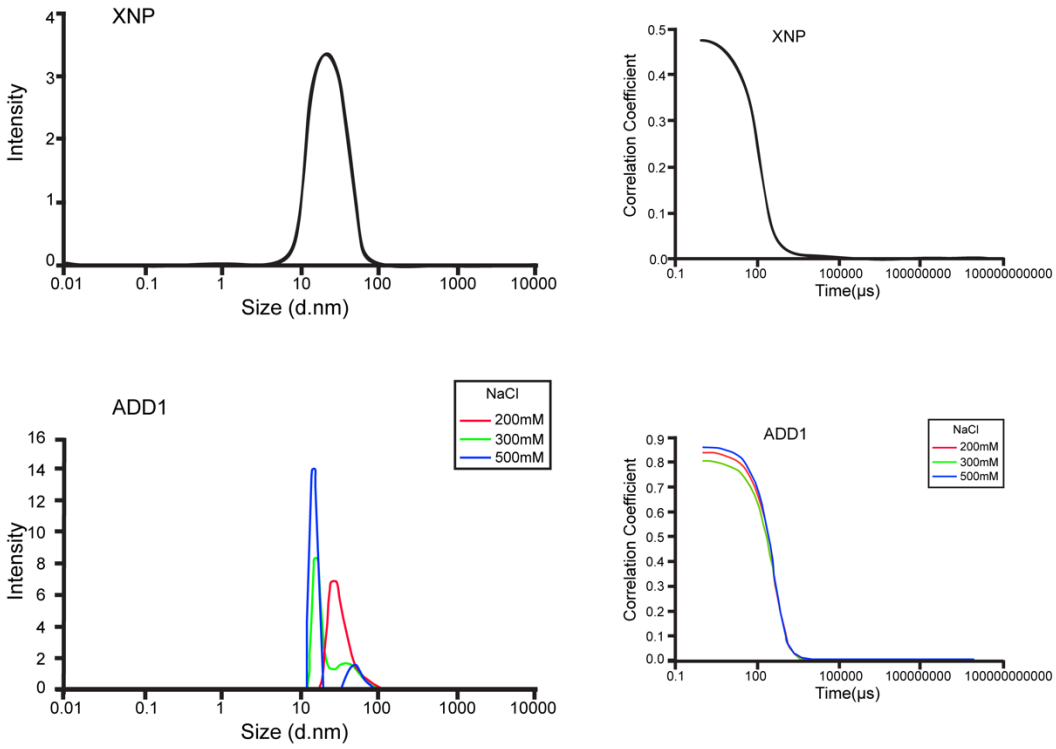


Figure 6–6 DLS Intensity plots for XNP and ADD1 purified protein

Dynamic light scattering (DLS) of recombinant full-length HIS-XNP-RA and ADD1-RA. Left: size distribution by mass plot showing intensity peaks. Right: raw correlation data. For ADD1-RA three salt concentrations are shown to highlight changes in the size distribution profile in buffer containing 200mM, 300mM and 500mM NaCl .

6.3.5 Identification of post-translational modifications on XNP and ADD1

Proteins modified by PTMs can run at different mobilities than expected based on their calculated molecular weight when unmodified, due to changes in protein size or structure influenced by folding or aggregation. I hoped, that by using mass spectrometry, I might be able to identify PTMs which could be responsible for the SDS denaturation resistant HIS-XNP-RA bands and modifications to explain the apparent size of ADD1-RA. The protein containing bands were excised from SDS-PAGE gels, washed, reduced and alkylated followed by in-gel tryptic digest and peptides were precipitated for identification by mass spectrometry. The peptides identified were scanned for post-translational modifications (PTMs) which were sustained through the purification procedure. 714 peptides mapping to ADD1-RA were identified in the purified ADD1-RA sample and for HIS-XNP-RA 391 peptides were identified. This amounted to 69% and 79% coverage of the proteins respectively (Figure 6–7). Post-translational modifications were concentrated in the N-terminal region of HIS-XNP-RA (Figure 6–7A). Ubiquitination on HIS-XNP-RA at site K181 was identified. For ADD1-RA protein a QQTGG(K) SUMOylation site was identified, in exon 4 of ADD-RA, which is only present in this ADD1 isoform, the longest isoform.

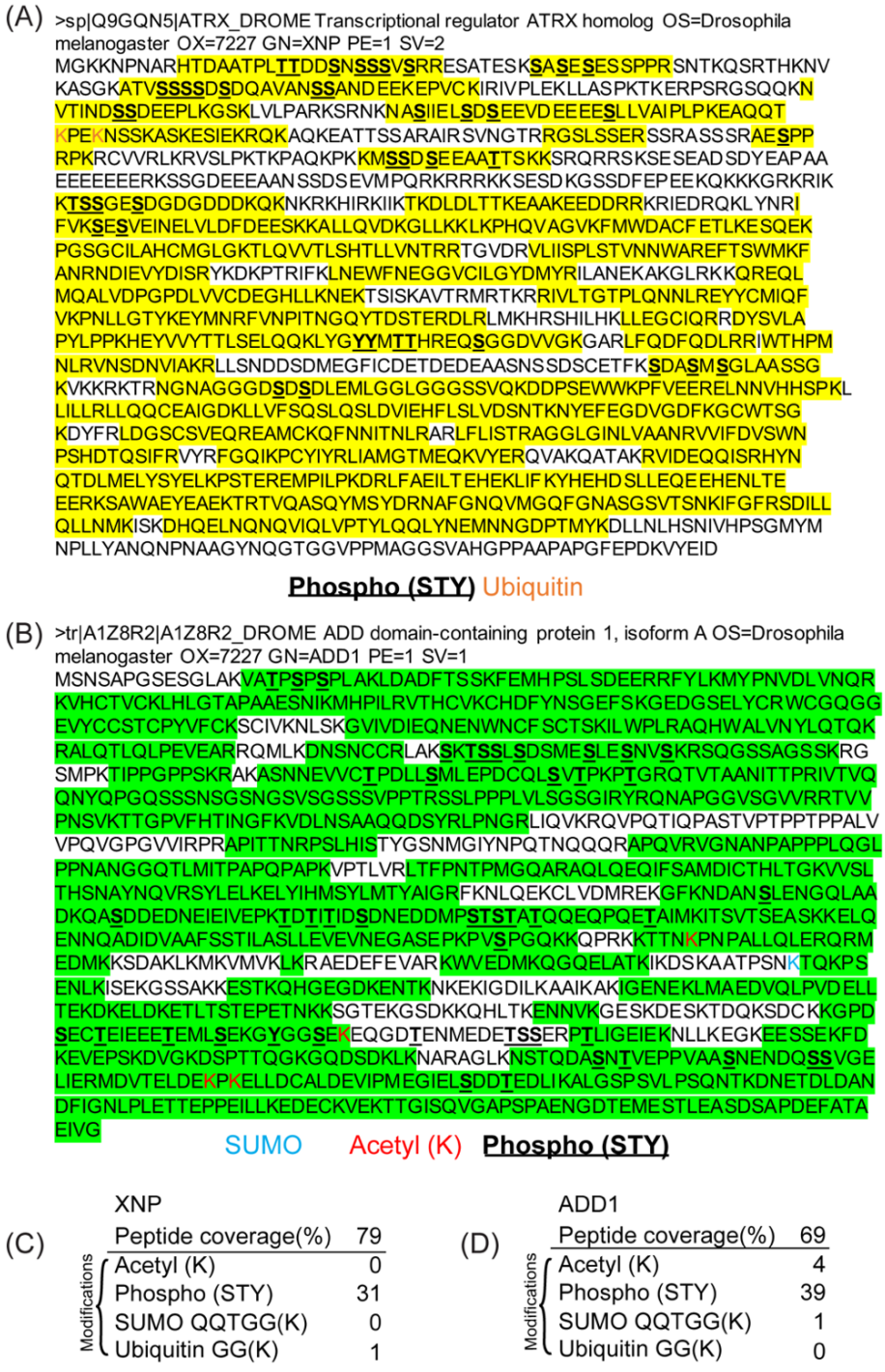


Figure 6–7 XNP and ADD1 post-translational modifications

(A and B) Amino acid sequences of ADD1-RA and XNP-RA, coloured sequences indicate peptide coverage from mass spectrometry of HIS-XNP-RA or ADD1-RA protein. The sites of PTMs identified by mass spectrometry are indicated **(C and D)** post translational modifications from (A and B) are summarised in table form.

6.3.6 Generation and testing of antibodies generated against XNP and ADD1

I generated polyclonal antibodies against the longest isoforms of ADD1 and XNP to enable polyclonal recognition of the entire proteins. One rabbit was immunised with 150µg ADD1-RA purified protein, given four boosts with a further 150µg ADD1-RA purified protein over a 90-day period, taking bleeds one week after each boost of antigen (see methods section 3.6.14), resulting in four bleeds including a pre-bleed. Pre-immune serum was compared to bleed one serum at day 32 and bleed two serum at day 60 (Figure 6–8). The XNP immunisation schedule, was started at a later date, and was terminated early due to closure of the company providing the immunisation schedule. Pre-immune serum was compared to bleed one serum and at the timing of the company's closure the rabbit was exsanguinated at the time point of taking bleed two.

To test the specificity of the antibodies in the rabbit bleed sera, western blots were performed. Extracts from ADD1 wild type, knock down and S2 cells transfected with pAc5.1-ADX-a-V5[296] to overexpress V5-ADD1 were run on SDS-PAGE followed by western blot and probed with anti-V5 antibody or anti-ADD1 bleed sera. A band of approximately 190kDA corresponding to the size of ADD1 was detected by both the anti-V5 antibody and the anti-ADD1 antibody bleed sera (Figure 6–8).

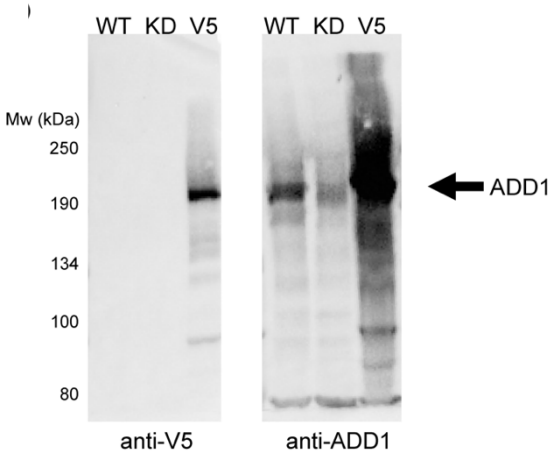


Figure 6–8 Testing of ADD1 serum from rabbit.

Western blots of cell lysates from wild-type, ADD1 sgRNA knockdown or transiently transfected V5-ADD1 S2 cells. Protein was quantified and equal amounts loaded per lane on 6% SDS-PAGE. Blots were probed with anti-V5 mouse antibody or anti-ADD1 rabbit bleed two serum.

To assess whether anti-XNP rabbit bleed one serum contained antibodies specific for XNP, the pre-immune and bleed sera diluted 1:500 was used in a western blot against exogenous FLAG-tagged XNP and wild-type S2 cell lysates (Figure 6–9). No protein was detected by probing with pre-immune serum, but the anti-XNP rabbit bleed one serum contained antibodies that recognised the same sized band as FLAG-XNP expressed from a plasmid. The predicted size of purified XNP is ~150kDa, but XNP from *Drosophila* lysate and S2 cells is detected as a band at 185kDa on western blot[139]. Antibodies in XNP rabbit bleed one serum can also detect endogenous *Drosophila* XNP protein on western blot.

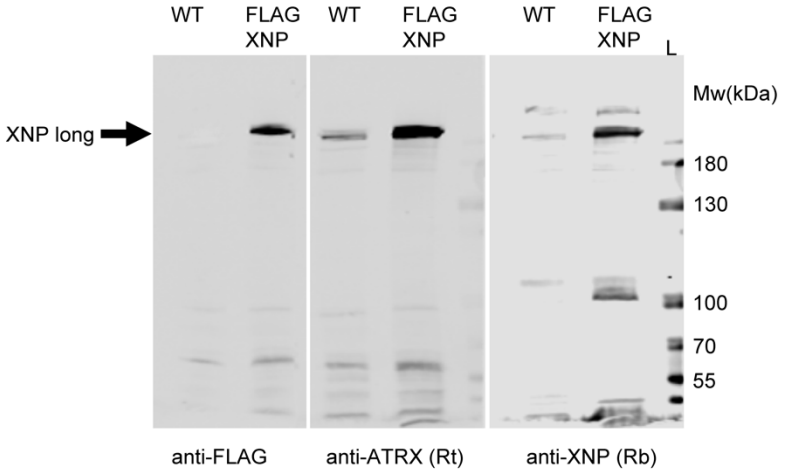


Figure 6–9 Testing of XNP serum from rabbit.

(A) Western blots of protein cell lysates from S2 cells either: wild-type (WT) or transiently transfected FLAG-XNP (FLAG-XNP). Protein was quantified and equal amounts loaded per lane on 8% SDS-PAGE. The blot was cut into 3 strips and probed with either: M2 anti-FLAG mouse antibody, anti-ATRX rat serum or anti-XNP rabbit bleed one serum (LHS) (diluted 1:500).

6.4 Summary

I aimed to characterise the interaction between ADD1, XNP, PIAS and SUMO *in vitro*. Efforts during purification to prevent aggregation and obtain properly folded ADD1 and XNP were unsuccessful as assessed by DLS. Therefore it was not possible to perform *in vitro* SUMOylation assays. Post-translational modifications on ADD1 and XNP were identified by mass spectrometry (Figure 6–7), confirming the presence of a Ubiquitin/SUMO modified XNP at the same site as identified previously, K181 (see Introduction 2.8). The misfolded protein was used as a source of antigen to generate antibodies. These were tested and found to recognise both the native and denatured form of ADD1 and XNP for use in western blot (Figure 6–8 and Figure 6–9), immunofluorescence (Figure 5–12) and immunoprecipitation (Chapter 7) to enhance our understanding of PIAS-XNP, PIAS-ADD1 and ADD1-XNP complexes, the results of which are discussed in results Chapter 7.

Many techniques including immunofluorescence and immunoprecipitation depend on having antibodies that are highly specific and selective for the protein of interest. An antibody able to recognise endogenous protein enables analysis of protein localisation and interactions in a physiological setting. Tagged proteins in cell lines were also used during this PhD study because detection of an epitope tagged protein using commercial antibodies provides independent confirmation of results.

6.4.1 Evaluation of the approach

6.4.1.1 Purified protein approach

Expression of large proteins is a challenge and was not possible for ADD1 and XNP at this stage. Properly folded proteins are essential for *in vitro* assays because protein conformation and structure influences protein function. ADD1 shows some salt sensitivity and increasing the salt concentration improved behaviour. Nevertheless, as the results indicated that both ADD1 and XNP aggregate, *in vitro* purification work was stopped.

Notably, ADD1 protein does not run at its predicted molecular weight on SDS-PAGE. MALDI-TOF analysis, knock-down and overexpression of ADD1-V5 allows me to conclude that the protein produced *in vitro* is ADD1.

The next step with the ADD1 and XNP protein purification would be to co-purify both proteins together from co-infected cells. It is possible that both proteins may need to be expressed simultaneously to stabilise each other and form a complex. If unsuccessful, truncation mutants of XNP could be expressed, focusing on the N-terminal sequence of XNP 1-223, only found in the long isoform of XNP and required for XNP localisation to heterochromatin[294] and would confirm the interaction observed in Lopez *et al.,* 2014[296].

Interestingly, ubiqutination/SUMOylation on XNP at site K181 was identified on protein expressed in Sf9 cells. This is in agreement with a previous finding investigating an N-terminal fragment of XNP (Heun lab unpublished) using a bacterial SUMOylation assay. A QQTGG(K) site was identified in exon 4 of ADD-RA, indicative of SUMOylation of ADD1 at site K164.

6.4.1.2 Testing antibodies

Due to early termination of the XNP immunisation programme of the rabbit, the affinity and titre of XNP antibodies may have been compromised, however, XNP rabbit bleed one serum was able to recognise endogenous XNP, demonstrating the presence of some antibodies in the serum with affinity for XNP protein.

This page is intentionally left blank

Chapter 7 Exploring the XNP-ADD1 interaction and their interaction with PIAS

7.1 Objective

In this chapter I used the XNP and ADD1 antibodies (generated as part of this PhD study, see section 6.3.6) for immunoprecipitation of XNP and ADD1 from S2 cells to identify whether ADD1, XNP, PIAS and SMT3 co-immunoprecipitate at endogenous protein levels.

7.2 Aims

- To identify endogenous XNP and ADD1 interactors
- To identify whether XNP and ADD1 interact with PIAS
- To identify common interactors between XNP and ADD1 in conditions where SUMOylation has been maintained
- To determine whether ADD1 is SUMOylated

7.3 Results

7.3.1 Detecting XNP SUMOylation

Immunoprecipitations were performed from S2 cell nuclear extracts using XNP rabbit Bleed 2 serum or pre-immune serum (IgG) coupled Protein G Dynabeads™ in the presence and absence of NEM. NEM prevents the removal of ubiquitin-like protein modifications (UBLs). If ubiquitin and SMT3 are covalently conjugated to XNP a size shift of approximately 9 kDa or 10 kDa respectively should be visible on western blot by probing with anti-XNP antibodies. XNP has been identified as a SUMOylated protein in *Drosophila* embryos in a mass spectrometry screen[171], but direct *in vivo* evidence is still lacking. A cell line expressing an endogenous tag on XNP would be ideal, however, attempts at generating an endogenously-tagged XNP S2 cell line were unsuccessful. Therefore, an S2 cell line with a stably retained pMT_ATRX_FLAG (Heun lab old nomenclature - ATRX refers to XNP) plasmid was used as a control to determine the specificity of my antibody against endogenous XNP. XNP can be inducibly expressed by the addition of copper sulphate from the pMT_ATRX_FLAG plasmid.

A low level of expression was induced from the pMT_ATRX_FLAG plasmid for 3 days prior to harvesting the cells for immunoprecipitation using the rabbit anti-XNP antibody. Immunoprecipitated material was run on SDS-PAGE and detected on western blot using either an M2 anti-FLAG mouse antibody or anti-XNP antibodies (Figure 7–1).

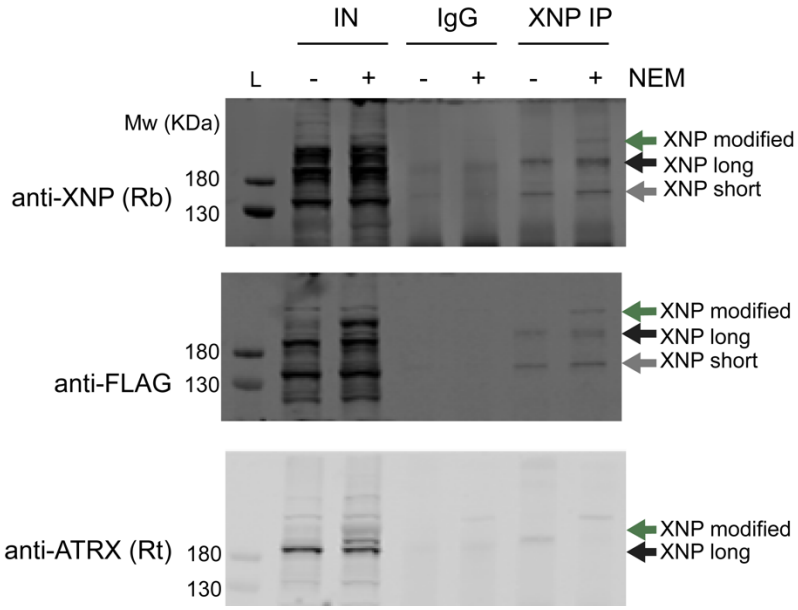


Figure 7–1 Immunoprecipitation of modified XNP

Nuclear extracts were made from wild-type S2 cells in buffers with or without NEM. Input (IN), IP with pre-immune serum (IgG) and XNP rabbit Bleed 2 serum (XNP IP) after washes with IP wash buffer containing 0.5M NaCl. were separated on 6% SDS-PAGE followed by western blot probed with anti-XNP rabbit (Rb) Bleed 2 serum (top panel), M2 anti-FLAG mouse antibody (middle panel) and *Drosophila* anti-ATRX rat (Rt) antibody (bottom panel). IP was carried out in the absence (-) or presence (+) or NEM detecting an extra band (green arrow) in the presence of NEM. The black arrow marks the expected size of the long isoform of XNP and the grey arrow the short isoform of XNP.

Rat and rabbit anti-XNP antibodies had different specificities and limits of detection. All antibodies recognise the long isoform of XNP, but the rabbit antibody also detects a slower migrating band, seen at a slightly higher molecular weight above the long isoform of XNP on a western blot, which was not detected with the anti-ATRX rat antibody (Figure 7–1). This additional band is visible for endogenous and FLAG-XNP extracts made in the presence of NEM (labelled XNP modified in Figure 7–1). Therefore it appears that XNP is post-translationally modified in S2 cells. Attempts to verify the identity of this larger band were unsuccessful, as it is difficult to detect endogenous UBLs on substrate proteins, as at any one time only a small proportion of the total protein is modified and the SMT3 and ubiquitin antibodies were not sensitive enough to detect any protein from this immunoprecipitated material.

7.3.2 Mass spectrometry of affinity purified XNP-associated complexes from S2 cells

The IP efficiency of XNP rabbit Bleed 2 serum was not optimal. Despite this, an IP-western blot using XNP rabbit Bleed 2 serum and probed with HP1a antibody (an interaction partner of XNP[139,299]) confirmed that this serum is able to co-immunoprecipitate HP1a-XNP-complexes (Figure 7–4). Bands absent from IgG control lanes were visible on SDS-PAGE silver stain (Appendix Figure 8–3).

To identify proteins that interact with XNP and modified XNP, in the absence and presence of NEM respectively, immunoprecipitation followed by mass spectrometry (IP-MS) was performed. Volcano plots were generated using Perseus to visualise the result of a Student's T-test comparing interactors identified in XNP-IP and IgG-IP in the absence and presence of NEM, after filtering for proteins identified in at least two out of three replicates and identified by two or more peptides per protein (Figure 7–2).

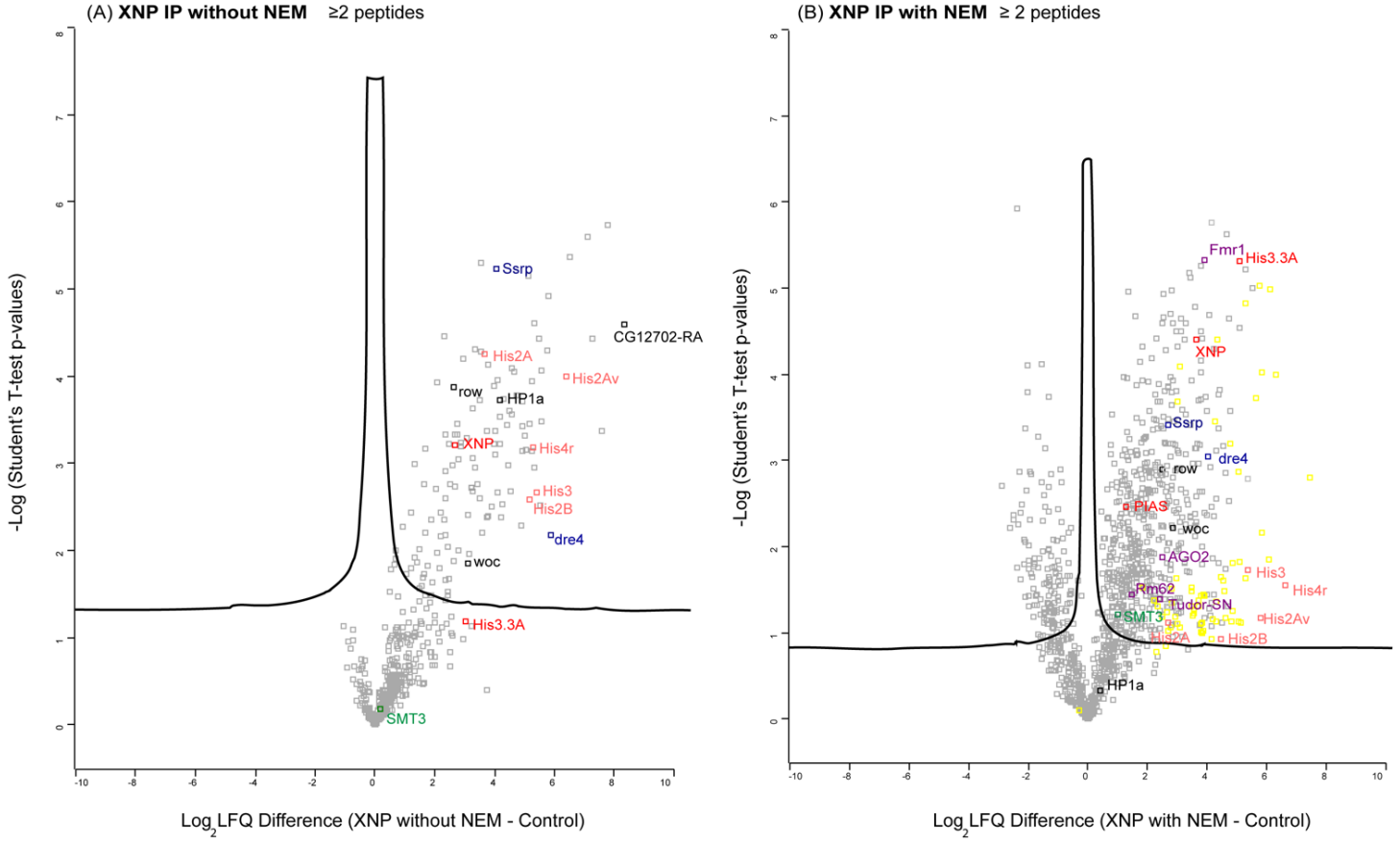

(A) XNP IP without NEM ≥2 peptides
-Log (Student's T-test p-values)
Log2 LFQ Difference (XNP without NEM - Control)
Ssrp
CG12702-RA
His2A
His2Av
row
HP1a
XNP
His4r
His3
His2B
dre4
woc
His3.3A
SMT3
(B) XNP IP with NEM ≥ 2 peptides
-Log (Student's T-test p-values)
Log2 LFQ Difference (XNP with NEM - Control)
Fmr1
His3.3A
XNP
Ssrp
dre4
row
PIAS
woc
AGO2
His3
His4r
Rm62
Tudor-SN
SMT3
His2Av
His2A
His2B
HP1a

Figure 7–2 Volcano plot of XNP protein interactors in S2 cells in the absence and presence of NEM

Immunoprecipitation (IP) using *Drosophila* anti-XNP rabbit Bleed 2 serum (XNP) or rabbit pre-immune serum (Control) IgG coupled to Protein A Dynabeads™ was performed from S2 cell lines. IP samples were prepared as described in methods and analysed by mass spectrometry. XNP interactors in conditions without NEM (**A)** and with NEM **(B)** are plotted. Each dot represents one protein. To determine the XNP interactome, a MaxQuant label-free quantitation pipeline was used. Student's T-test was performed on $\log_2$ transformed LFQ values from three control and three XNP IPs after filtering for proteins identified by two or more peptides and found in at least two of three replicates. T-test difference ratios (x-axis) were plotted against the negative logarithmic p-value of the T-test (y-axis). Named proteins in red are proteins of interest to this PhD study including XNP, H3.3 and PIAS. Proteins in purple are linked to RNAi silencing pathways. Proteins in black are consistent XNP interactors. SMT3 is in green. Proteins in yellow are 40S and 60S ribosomal proteins, which are likely to be non-specific interactors. The confidence line was generated using a permutation-based FDR value of 0.05 and a minimal fold change value of 0.1.

124 proteins were significant enriched in -NEM conditions and 608 proteins were significantly enriched in +NEM conditions (Figure 7–2). In the XNP IP without NEM the histone variant H3.3 is not significantly enriched, whereas other histone proteins are some of the most significant interactors (Figure 7–2). However, in the presence of NEM, H3.3 is more than four-fold enriched and has a higher -log10 significance value compared to other histone proteins (Figure 7–2). woc and row were identified by 20 or more peptides in XNP IP-MS (Figure 7–2) and were also found in in the PIAS IP-MS from S2 cells (Figure 4–16).

Published interactions of XNP were found in this study, these include the Exocyst complex component 3, Sec6[338] (identified by 4 peptides) and components of the Facilitates Chromatin Transcription (FACT) complex: Ssrp (identified by 4 peptides) and dre4 (identified by more than 20 peptides). However DNA replication-related element factor[339] (Dref), which is involved in insulator function, chromatin remodelling, and telomere maintenance, was not detected. Also, ADD1 was not identified in either the presence or absence of NEM.

A comparison of the XNP interactome in conditions with and without NEM found Tudor-SN, Fmr1 and Su(var)3-9 were only present in + NEM conditions (Figure 7–3). HP1a, woc, row and Ssrp were detected by LFQ values that differed less than two-fold in the absence and presence of NEM. HP1a was detected at higher levels in XNP IP in conditions without NEM (Figure 7–3).

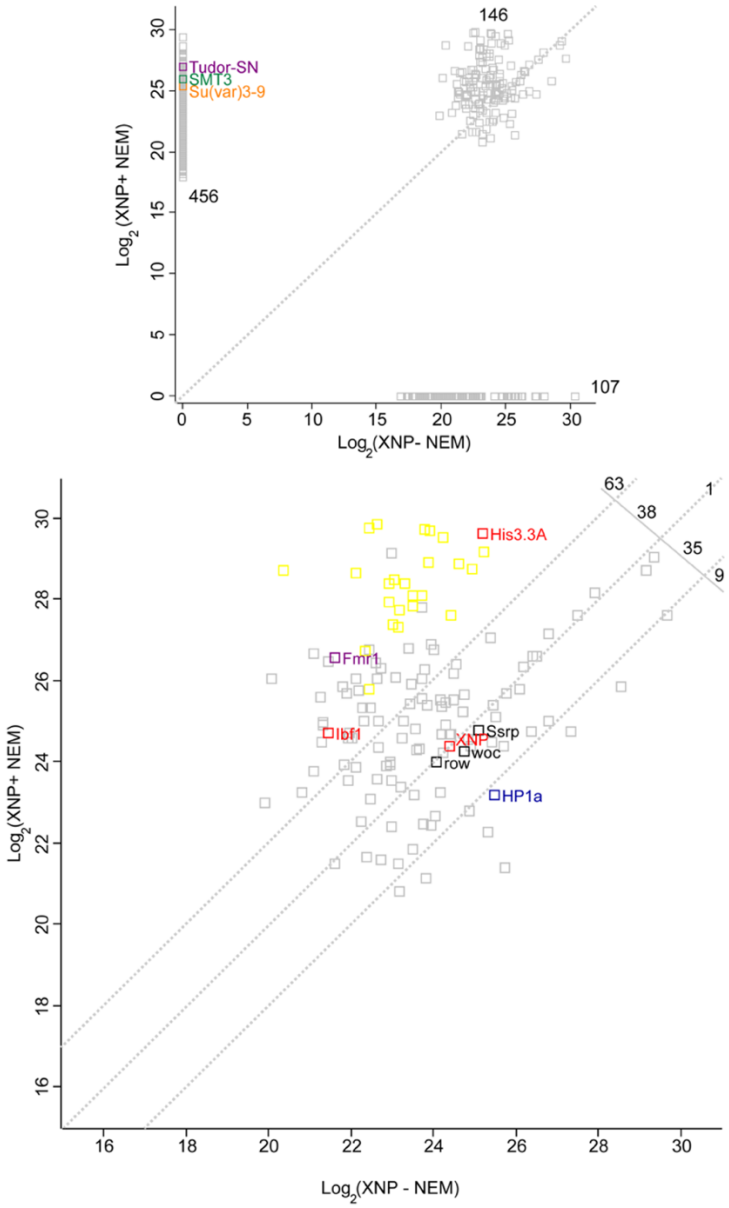


Figure 7–3 Scatter plot comparing XNP protein interactors in S2 cells in the absence and presence of NEM

Mean log_2 transformed LFQ values calculated from three biological replicates, normalised to XNP, from XNP IP in +NEM and -NEM conditions. Each dot represents one protein. The top plot shows all proteins and the bottom plot shows proteins found only in both conditions. Named proteins in red are proteins of interest to this PhD study including PIAS, XNP and H3.3. Proteins in purple are linked to RNAi silencing pathways. Proteins in black are consistent XNP interactors. SMT3 is in green. Yellow dots are ribosomal proteins which are likely to be non-specific interactors. Diagonal lines mark two log_2 fold change from zero.

7.3.3 XNP-ADD1 interaction with PIAS

Lopez Falcon 2014[296] identified an interaction between XNP and ADD1 and mapped this interaction to the first 266 amino acids of XNP (only present in the long isoform of XNP). I performed IP of endogenous XNP using anti-XNP rabbit bleed serum and IP of endogenous ADD1 using anti-ADD1 rabbit bleed 3 serum. The samples were analysed on SDS-PAGE followed by western blot. This showed PIAS and HP1a co-immunoprecipitate with ADD1 and XNP in the presence of NEM (lanes 6 and 13 of Figure 7–4). This is a weak interaction (identified by running 1% input and 100% IP). Although the anti-XNP rabbit antibody is more sensitive, the anti-ATRX rat antibody was used for co-probing blots with LICOR antibodies so all immunoprecipitated material could be loaded.

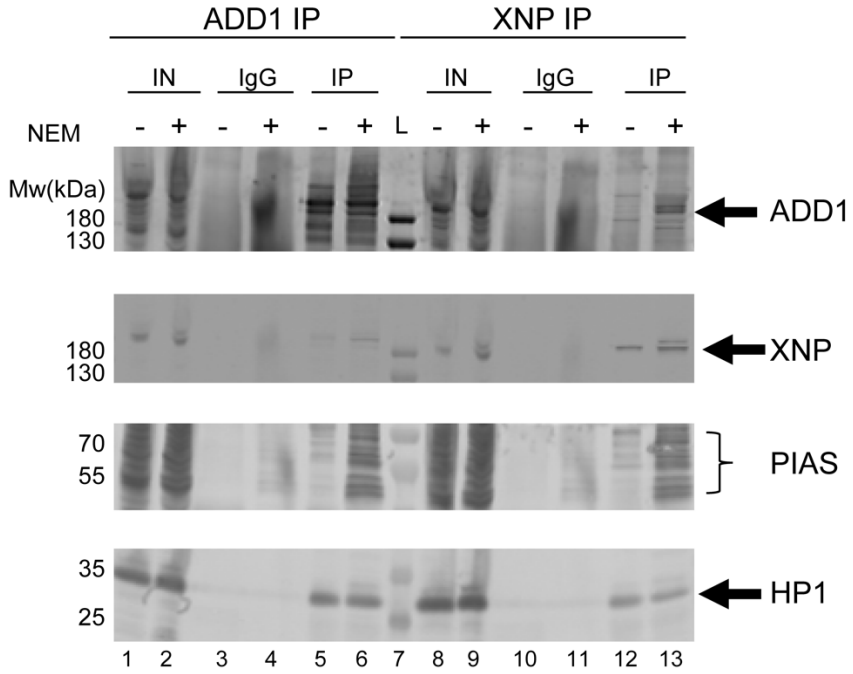


Figure 7–4 Co-immunoprecipitation of endogenous ADD1 and XNP

Nuclear extracts made from wild-type Schneider S2 cells in the absence and presence of NEM. The input (IN) and Immunoprecipitation (IP) using rabbit pre-immune serum (C), anti-ADD1 rabbit Bleed 3 serum (ADD1 IP) or anti-XNP rabbit Bleed 2 serum (XNP IP) coupled to Protein A Dynabeads™ were run on 4-20% SDS-PAGE and analysed by western blot. The western blot was cut and probed with anti-ADD1 rabbit bleed 3 serum (top panel), anti-ATRX rat antibody (second panel), anti-PIAS sheep antibody (third panel) or anti-HP1a mouse antibody (bottom panel). L=ladder

For independent validation of the interaction between ADD1 and XNP being dependent on SUMOylation, wild type and HA-ADD1 S2 cell lines (generated in this study, Appendix Figure 8–11) were both transfected with the pMT_FLAG_ATRX plasmid (from which XNP can be inducibly expressed when copper sulphate is added). S2 cell lines were generated with this plasmid and maintained under hygromycin selection, resulting in stable HA-ADD1; FLAG-XNP S2 cells and FLAG-XNP S2 cells. When XNP is expressed from an inducible promoter a larger pool is available in the soluble fraction. Cell lysate from wild-type, HA-ADD1 FLAG-XNP S2 cells and FLAG-XNP S2 cells were subject to FLAG pulldown using M2 resin. HA-ADD1 co-immunoprecipitated with FLAG-XNP only in the presence of NEM (Figure 7–5).

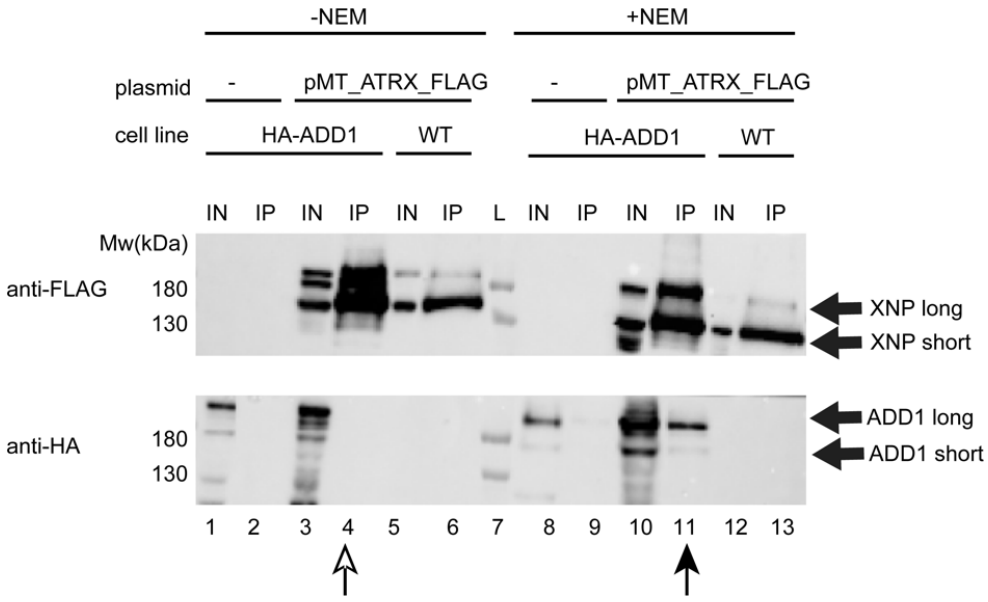


Figure 7–5 XNP co-immunoprecipitates with ADD1 in the presence of NEM.
HA-ADD1 S2 cells and wild-type S2 cells that maintain the pMT_FLAG_ATRX plasmid under hygromycin selection were generated, resulting in stable FLAG-XNP HA-ADD1 and stable FLAG-XNP cell lines. Expression of FLAG-XNP was induced from the pMT_FLAG_ATRX plasmid by addition of 200μM $CuSO_4$ for 3 days prior to harvesting. Nuclear extracts were made and subjected to FLAG pulldown using anti-FLAG M2 resin. HA-ADD1 S2 cells lysate was used as a negative control. The inputs (IN) and immunoprecipitations (IP) were analysed on 6% SDS-PAGE followed by western blot probed with anti-FLAG rabbit antibody (top panel) or anti-HA rat antibody (bottom panel). IP was carried out in the absence (-) or presence (+) or NEM detecting co-immunoprecipitation of ADD1 with XNP (lane 11, filled arrow) in the presence of NEM, but not in the absence of NEM (lane 4, hollow arrow).

7.3.4 Mass spectrometry of affinity purified ADD1-associated complexes from S2 cells

I immunoprecipitated endogenous ADD1 with the aim of identifying protein interactors in common with XNP in the presence and absence of NEM. Using affinity capture mass spectrometry, Alekseyenko *et al.,* 2014[299] identified: CG14438, HP1b, Su(var)205 (HP1a), Su(var)2-HP2, Eggless, CG3680, CG6791, CG1910, bon and mod(mdg4) as interactors of ADD1, but not XNP[299]. However this was performed without NEM. Therefore if an interaction occurs through maintenance of UBL PTMs, it will not be detected. The same conditions as used for XNP IP-MS were used for ADD1 IP-MS, to address the possibility that ADD1 and XNP interactions partners change in the presence of NEM. IP efficiency and specificity using anti-ADD1 rabbit bleed 3 serum compared to pre-immune serum was confirmed by silver stain (Appendix Figure 8–3).

Volcano plots were generated using Perseus to visualise the results of a Student's T-test comparing proteins identified in IP using anti-ADD1 rabbit bleed 3 serum to pre-immune serum IP in the absence and presence of NEM (Figure 7–6).

This page is intentionally left blank

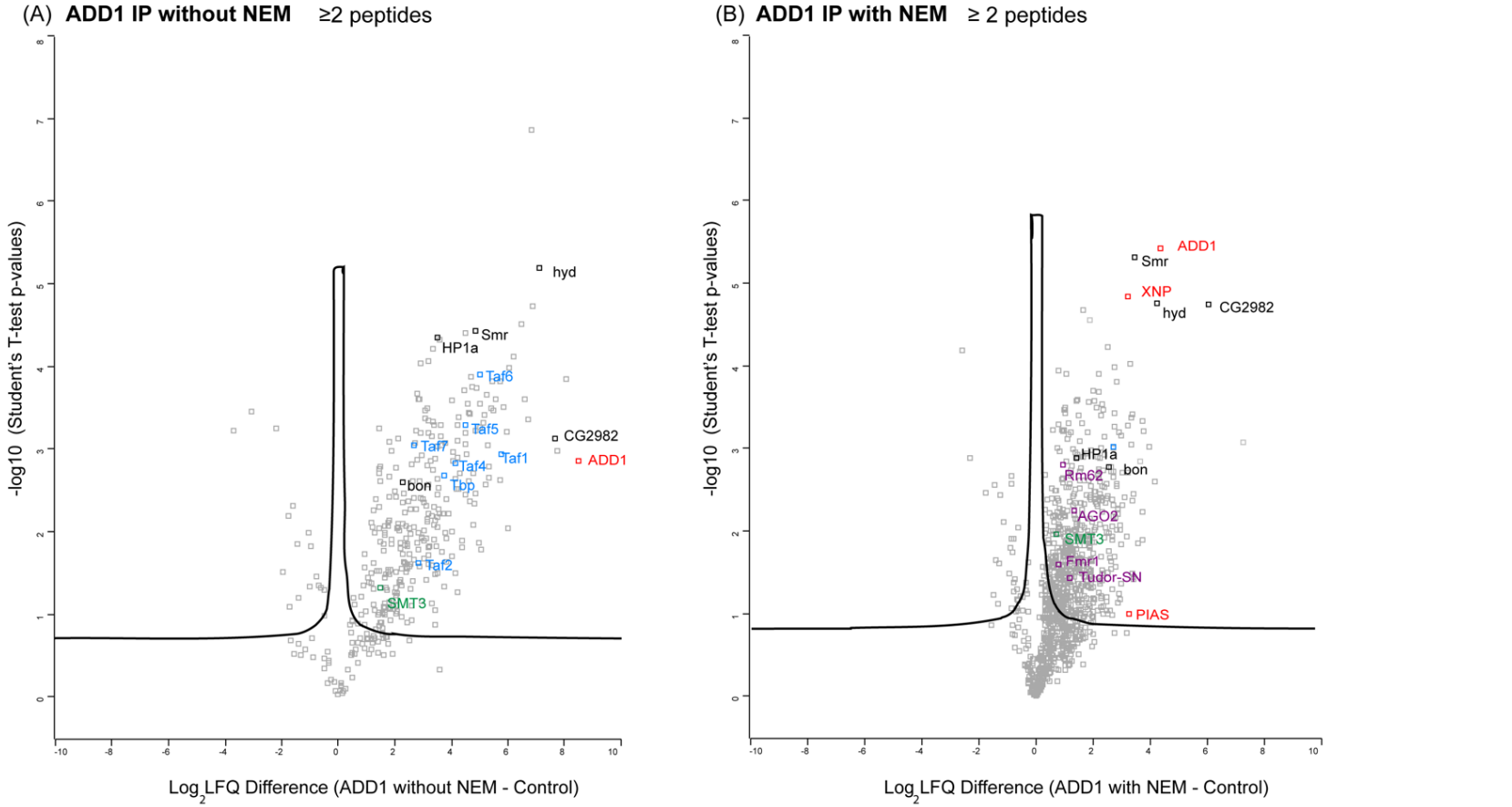

(A) ADD1 IP without NEM ≥2 peptides
-log10 (Student's T-test p-values)
Log2LFQ Difference (ADD1 without NEM - Control)
hyd
Smr
HP1a
Taf6
Taf5
Taf7
Taf1
Taf4
CG2982
ADD1
bon
Tbp
Taf2
SMT3
(B) ADD1 IP with NEM ≥ 2 peptides
-log10 (Student's T-test p-values)
Log2LFQ Difference (ADD1 with NEM - Control)
ADD1
Smr
XNP
hyd
CG2982
HP1a
bon
Rm62
AGO2
SMT3
Fmr1
Tudor-SN
PIAS

Figure 7–6 Volcano plot of ADD1 protein interactors in S2 cells in the absence and presence of NEM

Immunoprecipitation (IP) using *Drosophila* anti-ADD1 rabbit Bleed 3 serum (XNP) or rabbit pre-immune serum (Control) IgG coupled to Protein A Dynabeads™ was performed from S2 cell lines. IP samples were prepared as described in methods and analysed by mass spectrometry. ADD1 interactors in conditions without NEM (**A)** and with NEM **(B)** are plotted. Each dot represents one protein. To determine the ADD1 interactome, a MaxQuant label-free quantitation pipeline was used. Student's T-test was performed on $\log_2$ transformed LFQ values from three control and three ADD1 IPs after filtering for proteins identified by two or more peptides and found in at least two of three replicates. T-test difference ratios (x-axis) were plotted against the negative logarithmic p-value of the T-test (y-axis). Named proteins in red are proteins of interest to this PhD study including PIAS, ADD1 and XNP. Proteins in purple are linked to RNAi silencing pathways. Proteins in black are consistent ADD1 interactors. Proteins in blue are transcription activation proteins (Tafs). SMT3 is in green. The confidence line was generated using a permutation-based FDR value of 0.05 and a minimal fold change value of 0.1.

The E3 ubiquitin ligase hyd, SANT domain protein SMRTER (Smr) and Bifunctional lysine-specific demethylase and histidyl-hydroxylase NO66 (CG2982) were identified as interactors in both the absence and presence of NEM (Figure 7–6). In both conditions ADD1 co-immunoprecipitated with HP1a, bon and Tafs (Figure 7–6).

A comparison of the ADD1 interactome in conditions with and without NEM recovered an additional 755 proteins in the presence of NEM (Figure 7–7). Interactors identified in ADD1 IP-MS in both conditions, such as woc and row, were detected by higher LFQ values in NEM conditions, suggesting SUMOylation influences ADD1's interaction partners or interaction strength. Transcriptional activator proteins (Tafs), hyd and HP1a were detected by LFQ values that differed less than two-fold in the absence and presence of NEM. This supports that their interaction with ADD1 does not change in the presence and absence of SUMOylation (Figure 7–7).

XNP was detected only in the presence of NEM suggesting SUMOylation of either XNP, ADD1 or intermediate interaction partners is needed for ADD1 and XNP to interact. Other interactors only present in NEM IP conditions include PIAS, Tudor-SN and Fmr1 (Figure 7–7).

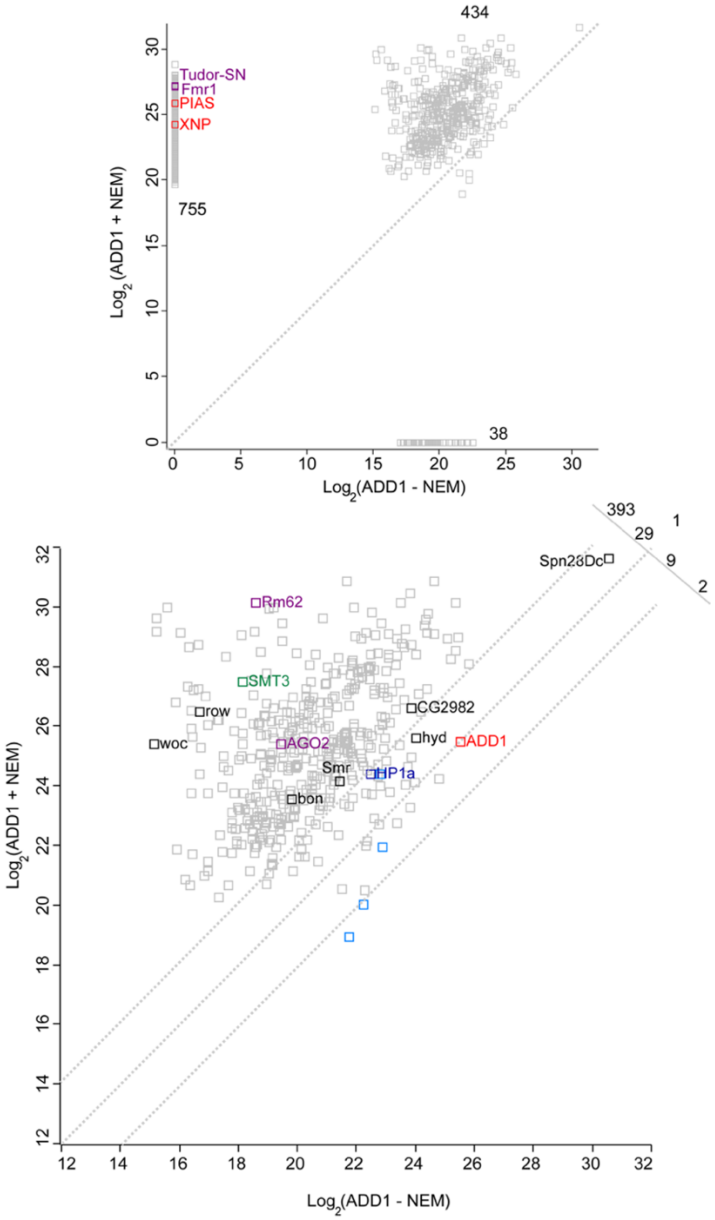


Figure 7–7 Scatter plot comparing ADD1 protein interactors in S2 cells in the absence and presence of NEM

Mean log_2 transformed LFQ values calculated from three biological replicates, normalised to ADD1, from ADD1 IP in +NEM and -NEM conditions. Each dot represents one protein. The top plot shows all proteins and the bottom plot shows proteins found only in both conditions. Proteins in red are proteins of interest to this PhD study including PIAS, ADD1 and XNP. Proteins in purple are linked to RNAi silencing pathways. Proteins in black are consistent ADD1 interactors. Proteins in blue are transcription activation proteins (Tafs). SMT3 is in green. Diagonal lines mark two log_2 fold change from zero.

7.3.5 Interactors in common between ADD1 and XNP IP-MS lists

In IP-MS analysis, the following criteria were applied to identify shared interactors:

- proteins immunoprecipitated by ADD1 and XNP in the presence of NEM
- identified by two or more peptides
- in at least two out of three replicates
- not identified by two or more peptides in pre-immune serum controls.

These are shown in Table 16. The abundance of the different interactors cannot be directly compared due to differential recovery from different IP targets.

Shared interactors of ADD1 and XNP were used to query the protein interaction database IntAct (https://www.ebi.ac.uk/intact/) to build a network of interactors, with the aim of identifying proteins with common roles in particular cellular processes. The software Cytoscape was used to visualise the network as a node and edge diagram. The information in this database is derived from literature curation. Of the 48 proteins, only 25 appeared in the IntAct database. Fundamentally, XNP, ADD1 and PIAS could not be linked to interactors. However, useful information could still be obtained, most notably that rhea, Rbcn-3A, cora and Pescadillo homologs are all enriched in the nervous system, where human ATRX mutations cause a mental retardation phenotype. Identification of these proteins supports a role for the ADD1 and XNP proteins acting in common and conserved pathways. PIAS is found as a co-interactor of both ADD1 and XNP only when SUMOylation is maintained in the presence of NEM. This indicates that SUMOylation of these proteins and this complex may be essential for their role and mechanism in ATRX's mental retardation phenotype.

14 out of the 48 common interactors are involved in chromosome organisation: Brahma and Mi-2 are both involved in ATP-dependent chromatin remodelling and Brahma in H3K27 acetylation; Ebi is involved in histone deacetylation; Maleless is involved in DNA duplex unwinding and heterochromatin assembly; JIL-1 is involved in histone phosphorylation; Cap-G, PIAS and Mi-2 are involved in chromosome condensation. In *Drosophila* Rif1 is involved in maintenance of chromatin silencing at the telomere. In addition to Rif1, mre11 and JIL-1 have roles in telomere maintenance.

This page is intentionally left blank

Table 16 Proteins in common in ADD1 and XNP IP-MS in the presence of NEM

Proteins identified in two out of three replicates by two or more peptides and not identified in the pre-immune serum negative controls by more than one peptide are listed.

Majority protein IDs	Protein names	Function	Median peptide count ADD1	Median peptide count XNP
A0A0B4KHR8	Ankyrin repeat and KH domain-containing protein mask	scaffolding	17	37
Q9VSL8	Rhea, isoform G	adaptor	11	30
M9PB68	Protein purity of essence	E3 ubiquitin ligase	6	30
A0A0B4LFW6	Receptor mediated endocytosis 8, isoform A	heat shock	5	22
A0A0B4KEY4	Patronin, isoform I	scaffolding	15	21
P51592	Hyperplastic discs, isoform B	E3 ubiquitin ligase	20	18
Q9W425	Rabconnectin-3A	Notch signalling	10	18
A0A0B4LHR3	Transcriptional regulator ATRX homolog	chromatin remodelling	4	18
A0A0B4K7T7	Rho guanine nucleotide exchange factor 2, isoform D	enzyme	8	16
Q9VVA4	Uncharacterized protein	enzyme	8	16
P25439	Brahma, isoform E	chromatin remodelling	12	13
Q9V8R9	Coracle, isoform F	scaffolding	7	13
X2JBC1	Regulator of nonsense transcripts 1 homolog	mRNA NMD	6	13
A0A0B4LEK8	Sec24AB ortholog, isoform A	transport	9	12

				XNP-ADD1
Q9W543	Rabconnectin-3B, isoform A	Notch signalling	8	12
M9PBG5	Enhancer of bithorax (Ebi)	chromatin remodelling	12	11
Q9V466	Nuclear pore complex protein Nup107	DSBR	11	11
Q7KVD1	Trio, isoform A	scaffolding	10	11
Q8SWR8	Ataxin-2, isoform D	miRNA pathways	6	10
M9PIA6	Chromodomain-helicase-DNA-binding protein Mi-2 homolog	chromatin remodelling	10	9
Q9VP27	Golgi complex-localized glycoprotein 1, isoform B	enzyme	7	9
Q7KQM6	GIGYF family protein CG11148	autophagy	6	9
Q86BR6	Autophagy-related	autophagy	4	9
Q9VCN1	DNA polymerase epsilon catalytic subunit	DNA replication	11	8
Q9VHI1	Hyrax	signalling	9	8
A1Z987	Chromosome associated protein G, isoform G (Cap-G)	condensin I complex	8	8
Q961C3	Uncharacterized protein	N/A	6	8
Q9VJH2	Aspartyl-tRNA synthetase, mitochondrial, isoform A	enzyme	12	7
Q9XZ34	Telomere-associated protein RIF1	DNA replication	10	7
A0A0C4DHA1	Chondrocyte-derived ezrin-like domain containing protein, isoform E	scaffolding	7	7
Q7K0D8	Nuclear pore complex protein Nup50	scaffolding	6	7
M9PGQ6	Puff-specific protein Bx42	Notch signalling	6	7
Q9W0M7	Splicing factor 3B subunit 3	mRNA processing	5	7
Q95RJ9	F-box-like/WD repeat-containing protein ebi	Notch signalling	5	7
O18413	26S proteasome regulatory subunit 8	protein degradation	5	7

Q9VTE6	IP14658p	ribosome biogenesis	4	7
O46048	Enhanced adult sensory	chromosome associated	12	6
M9PI05	Chromosomal serine/threonine-protein kinase JIL-1	chromatin structure	11	6
Q9XYZ4	Double-strand break repair protein (mre11)	DSBR	6	6
Q9VWD9	LD38919p	protein degradation	5	6
Q9VIW7	tRNA-splicing ligase RtcB homolog	mRNA processing	4	6
P24785	Dosage compensation regulator, maleless	chromatin associated	9	5
Q9VWK5	GH15225p/Ulp1	protein deSUMOylation	9	5
A1Z7P7	Suppressor of variegation 2-10, isoform M	SUMO E3 ligase	7	5
Q9VL96	Pescadillo homolog	ribosome biogenesis	5	5
Q9VBQ5	LD38433p	enzyme	5	5
M9NDW1	DIS	JNK pathway	6	4
A0A0B4LG05	Maternal protein tudor	Piwi-RNA pathway	5	4

7.3.6 SUMOylation modifications on ADD1 and XNP

Algorithms have been developed to predict where post translational modifications (PTMs) are likely to occur on proteins. Putative SUMOylation sites were determined by searching the full-length amino acid sequence of ADD1-RA (A1Z8R2) and XNP-RA (Q9GQN5) with three different PTM prediction tools: SUMOplot (http://www.abgent.com/sumoplot), GPS-SUMO (http://sumosp.biocuckoo.org/) and JASSA (http://www.jassa.fr). For ADD1-RA, two lysine residues were predicted as SUMOylation sites by GPS-SUMO whereas in SUMOplot 21 sites were identified. For XNP-RA, three SUMOylation sites were identified in GPS-SUMO and 12 in SUMOplot. There was no overlap in the predictions, with predicted SUMO and SIM sites not predicted to be in consensus sites.

Unlike XNP that has been identified as a SUMO substrate[171], it was not known whether ADD1 itself is SUMOylated. A higher molecular weight smear is visible above the main HA-ADD1 bands, characteristic of modified forms of ADD1 (Figure 7–8). ADD1 co-immunoprecipitates with SUMOylated species when NEM is present (Figure 7–8).

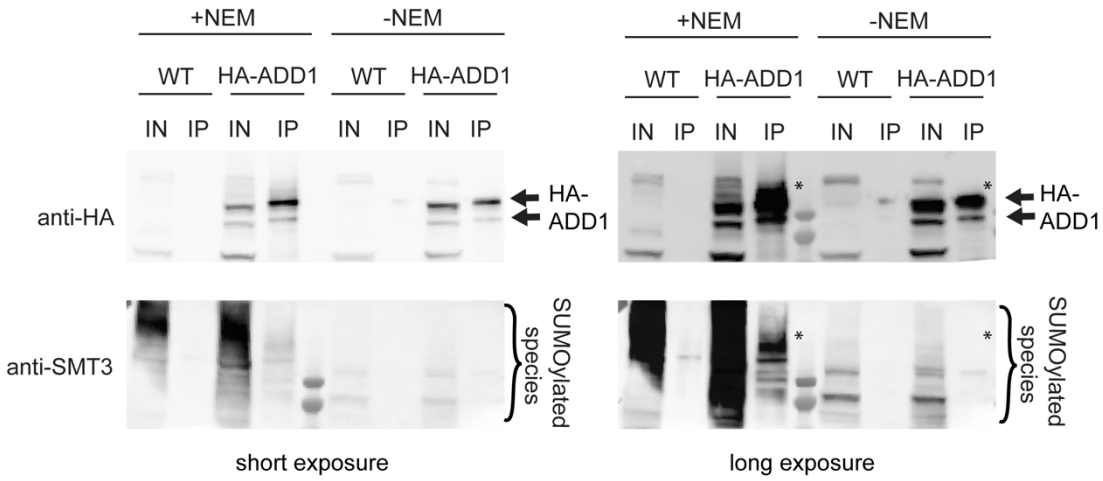


Figure 7–8 ADD1 co-immunoprecipitates with SUMOylated species.
Nuclear extracts were made from wild-type (WT) or HA-ADD1 S2 cells. The input (IN) and IP with anti-HA resin (IP) were run on 8% SDS-PAGE and analysed by western blot probing with anti-HA mouse antibody or anti-SMT3 rabbit antibody. SUMOylated species are marked by an asterisk in the longer exposure western blot.

A large number of *in vivo* validated SUMOylation sites do not occur on consensus sequences[176]. To determine where XNP and ADD1 are post-translationally modified when expressed at endogenous levels, I searched the MS/MS spectra from S2 cell and embryo IP-MS for SUMOylation sites. Despite having high overall peptide counts, many peptides that mapped to ADD1 and XNP had the same sequence. One QQTGG (K) site was identified in ADD1, mapping to exon 4 at the same site found in Sf9 expressed protein. The mass spectrum is shown for this peptide (Figure 7–9). No QQTGG (K) sites were identified in XNP.

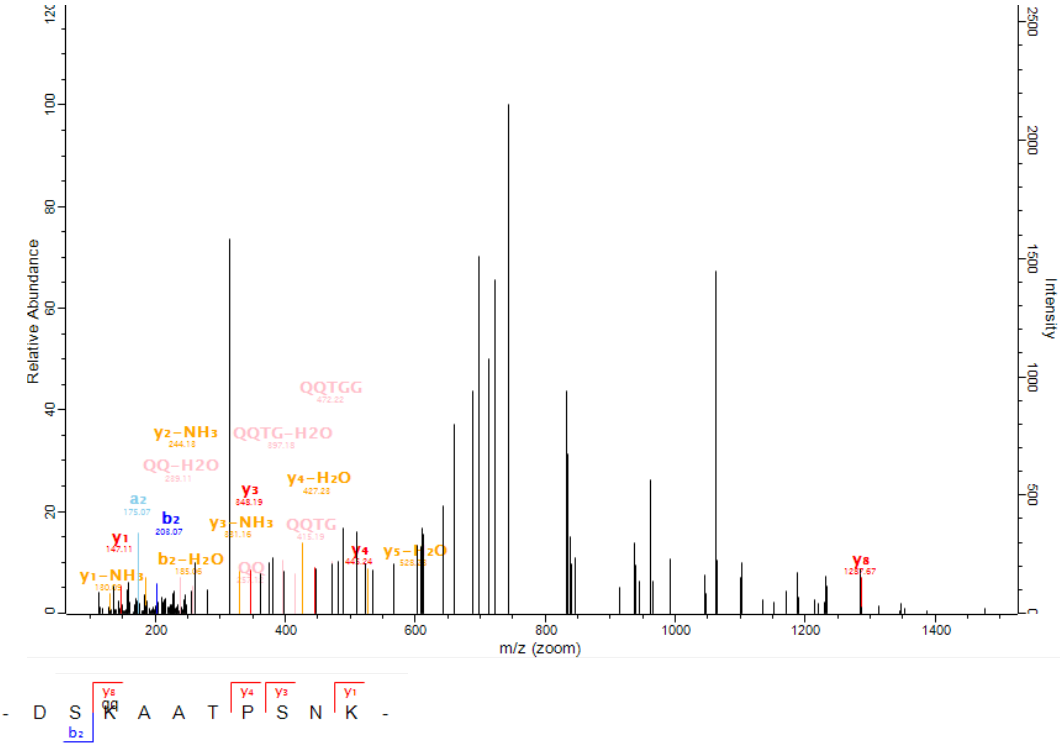


Figure 7–9 MS/MS spectrum of an ADD1 peptide modified by QQTGG(K)

Tandem mass spectrometry (MS/MS) spectrum of a QQTGG (K) site on a peptide identified in XNP IP-MS indicating SUMOylation of ADD1 lysine-146. The b and y ions are marked. This modification was identified once.

7.4 Summary

In this chapter I have demonstrated that the ADD1 and XNP antibodies could detect modified forms of ADD1 (Figure 7–8) and XNP (Figure 7–1) in the presence of NEM in S2 cells. Probing the immunoprecipitated material with anti-SMT3 antibodies could not confirm that SMT3 was covalently attached to XNP nor ADD1. We hypothesised that the interaction between ADD1 and XNP might be mediated by SUMOylation of XNP, promoted by PIAS. Mass spectrometry approaches to identify post-translational modifications on XNP and ADD1 had limited success due to low peptide coverage of the proteins. However, one SUMO site was identified on ADD1 at K-146 (Figure 7–9). Together, with previous data, this supports the likelihood of SUMO-related functions.

The XNP-HP1a interaction shown by Emelyanov *et al.,* 2010[139] and ADD1-HP1a by Alekseyenko *et al.,* 2014[299] most likely occurred in conditions where post-translational modifications had been lost. I hypothesise that ADD1 might only interact with modified XNP which is not very abundant in the cell. My IP-MS datasets comparing conditions where SUMOylation has been maintained opens up new interaction partners to explore.

7.4.1 XNP interactors

In mouse cells, recruitment of SSRP1 (a FACT complex component) to heterochromatin is dependent on ATRX[273], with which it can interact directly[340]. SSRP1 can be modified by ubiquitin-like modifications that have been identified through exogenous expression of SUMO in mouse cells[305]. FACT is an evolutionarily conserved histone chaperone that recognises the nucleosome and destabilises the interactions between H2A-H2B dimers and the H3-H4 tetramer[341]. H3.3 was also enriched in PIAS +NEM conditions. Dref, a published XNP interactor involved in chromatin organisation[339], was not detected in my XNP IP-MS. ADD1 was also not identified (Figure 7–2). This could be due to the conditions used for IP. As the anti-XNP antibody was extracted from an early serum (antibody production stopped at bleed 2 and rabbit was terminated due to closure of the company making it), it may only weakly recognise XNP, which may affect capture of low abundance interactors. In addition, washes in IP wash buffer containing 0.5M NaCl may have

caused loss of some XNP-complexes and/or interactors during the immunoprecipitation procedure potentially due to a weakly binding antibody.

7.4.2 An interaction between ADD1, XNP, HP1 and PIAS

It is possible that the ADD1 and XNP interaction is indirect or requires the proteins to be in a particular post-translationally modified state. This would explain why ADD1 and XNP interaction is only found in some experimental settings and not others[296.342], including the presence of NEM (Figure 7–6). HP1a is a possible intermediate binding partner. ADD1 interacts with HP1a[299] and XNP interacts with HP1a[299] and both were identified in ADD1 IP-MS (Figure 7–6) and XNP IP-MS (Figure 7–2) in this study and in a HP1a IP-MS[299]. However, whether ADD1 and XNP bind directly needs further clarification. ADD1, XNP and HP1a localise to telomeric repeat sequence, where PIAS is also present[322]. Alternatively, my IP-MS shows that PIAS interacts with XNP and ADD1. As PIAS is an E3 SUMO ligase, it is possible that PIAS can SUMOylate HP1a, ADD1 and/or XNP, affecting their recruitment to chromatin, maintenance at chromatin or the stability of interaction partners.

The ADD1-XNP interaction could be context and/or cell cycle dependent, meaning that to capture it, as a rare event, is challenging. Understanding the dynamics of XNP SUMOylation would help identify a time and/or cell type in which to explore this interaction further. There are no other SUMO E3 ligases enriched as interactors of XNP and ADD1 (Table 16).

7.4.3 ADD1 interactors

Endogenous ADD1 IP-MS identified the E3 ubiquitin ligase hyd, the SANT domain protein SMRTER (Smr) and the bifunctional lysine-specific demethylase and histidyl-hydroxylase NO66 (CG2982) (Figure 7–6). CG2982 demethylates H3K4me3 and H3K36me, possibly controlling the balance between heterochromatin and euchromatin.

Interestingly, ADD1 remains associated with chromatin at mitotic chromosomes after HP1a disassociation[303] and is found at the boundaries of heterochromatin and euchromatin[303]. Through interaction with other proteins and protein complexes

ADD1 could control the balance between heterochromatin and euchromatin. PIAS is also weakly enriched at pericentromeric heterochromatin and sub-telomeric regions in mitosis[143]. I hypothesise that PIAS mediated SUMOylation of ADD1 at the G2/M transition could influence HP1a re-association with chromatin after mitosis.

7.4.4 Evaluation of the approach

7.4.4.1 Sensitivity and performance of ADD1 and XNP antibodies in detecting post-translationally modified forms of ADD1 and XNP

The anti-XNP rabbit antibody generated in this study is more sensitive in its ability to detect denatured protein on western blot than the anti-ATRX rat antibody (Figure 7–1). Both XNP and ADD1 rabbit antibodies generated in this study enrich for their target protein from a complex protein lysate (Figure 7–4).

7.4.4.2 IP-MS with ADD1 and XNP serum

For non-affinity purified antibodies, the XNP and ADD1 IP-MS experiments worked remarkably well, isolating ADD1 and XNP-associated complexes with enrichment above background levels and isolation of known and predicted targets (Figure 7–6 and Figure 7–2).

7.4.4.3 Overexpression of XNP as a soluble source of modified XNP

XNP can interact with ADD1 in the presence of NEM (Figure 7–4). This, together with Figure 7–5, presents the question of what is different between +NEM and -NEM conditions. Perhaps both ADD1 and XNP need to be modified to interact or to interact with an intermediate that is preserved in +NEM conditions, which facilitates interaction or recruitment of the proteins to form part of a larger complex. To confirm that the ADD1-XNP interaction is mediated by SUMOylation, and is not itself direct, ADD1 and XNP would need to be produced recombinantly and used in direct binding assays *in vitro*, without the presence of other proteins or PTMs. This was not possible during this body of work due to not producing stable ADD1 and XNP *in vitro*. This was the subject of chapter 7 of this thesis. I would also hypothesize that modification or absence of modification results in different stabilities of XNP.

7.4.4.4 Identification of post-translational modifications

A SUMOylation site motif on ADD1 was identified by analysis of MS/MS spectra for this post-translational modification. However, SUMOylation is not motif specific and SMT3 can bind at different protein sites. Confirming SUMOylation at lysine-146 of ADD1 would require mutation of the SMT3 C-terminus to generate a T86K site (see section 2.3.8.1), which can be digested by trypsin, leaving a small peptide chain attached to ADD1 that can be detected by mass spectrometry. This approach could also confirm the location of SUMOylation sites on ADD1, XNP and other detected proteins.

The standard approach for identifying the function of SUMOylation on a target protein is to map the *bona fide* SUMO target sites on the substrate protein and generate SUMOylation-deficient mutants by site-directed mutagenesis. The QQTGG (K) SUMOylation site identified in exon 4 of ADD1-RA is novel (Figure 7–9) and was found on both Sf9 expressed protein and in IP-MS from endogenous S2 cells. This is on a non-canonical sequence that was not identified by prediction based software. However, E3 ligases can enable modification of non-consensus sites[176]. Although SUMOylation may be added at alternative sites, the combination of data to support modification of K-146 on ADD1 provides a starting point for addressing the role of ADD1 SUMOylation in a system with reduced ADD1 SUMOylation by mutating the SUMO site.

Chapter 8 Conclusions, Perspective and Future Direction

In this thesis I identified interaction partners of endogenous *Drosophila* PIAS in S2 cells and embryos. I have identified the general and developmental-specific interactions, to discover PIAS interactors that may be involved in heterochromatin formation. A number of exciting new interactions were discovered. These include proteins in RNA processing pathways and the Piwi-piRNA pathway and the histone methyltransferase Eggless, that is responsible for heterochromatin formation.

I attempted to purify full length XNP-RA and ADD1-RA protein to characterise their direct interaction, which is possibly mediated by SUMOylation. Although unsuccessful, this allowed generation of antibodies for further study. These antibodies allowed immunoprecipitation of XNP and ADD1 to identify interaction partners with and without retention of SUMOylation. This demonstrated that PIAS, HP1, ADD1 and XNP co-immunoprecipitate under conditions where SUMOylation is maintained (Figure 7–4).

8.1 PIAS self-oligomerises, but does not reside in PIAS bodies in S2 cells at endogenous levels

In the nucleus, only a small percentage of proteins are SUMOylated at any given time. In the presence of NEM, where SUMO conjugation to substrates is maintained, PIAS is present in SDS resistant structures (consistent with covalent modifications) visible on western blot (Figure 4–8 and Figure 4–10), but the functional relevance is not known. In 2001 Kumar Hari speculated that PIAS forms interactions with chromatin and that these may be stabilised by PIAS self-association and/or interactions with other scaffolding proteins. A secondary consequence of this might be to establish nuclear domains containing a high concentration of silencing protein[143]. Excess PIAS, when overexpressed, forms nuclear bodies through SUMO-SIM interactions[144] probably indicating loss of the normal equilibrium.

I hypothesise that PIAS, when SUMOylated, could form or bind insoluble structures in the nucleus through its scaffold attachment domain[147]. Alternatively SUMOylation and PIAS may aid phase separation to create local concentrations of target proteins. Formation of local concentrations through either phase separation (as seen with PIAS bodies in PIAS-V5-PE overexpression) or scaffolding could be important to PIAS's role at euchromatin-heterochromatin boundaries.

8.2 PIAS interacts with boundary elements and RNA Pol II transcription components

PIAS interacts with the boundary element associated factor (BEAF-32) (Figure 4–18) and both proteins share similar genome localisation patterns in the embryonic *Drosophila melanogaster* cell line Kc167[57]. PIAS is categorised as binding euchromatin, which has a high density of regulatory complexes, and localises to gene start sites[57].

PIAS interacts with transcription associated factors (Tafs) of the TFIID complex irrespective of NEM (Figure 4–17), indicating a SUMO-independent interaction. TFIID complex components are important for RNA Pol II transcription initiation. It is possible that PIAS-mediated SUMOylation of factors required to localise to TFIID is important for transcription regulation. Consistent with these observations, human PIAS is also involved in transcription and directly interacts with TATA binding protein (TBP)[310], suggesting that a role in transcription may be evolutionarily conserved.

8.3 PIAS interacts with nucleic acid binding proteins

Given that SUMO is involved in a large number of cellular processes (see sections 2.3.6-2.3.13), it is unsurprising that many proteins were found to interact with PIAS. Biochemical (salt fractionation and size exclusion chromatography) experiments demonstrated that PIAS is present in multiple complexes (Figure 4–2 and Figure 4–9). It is likely that PIAS isoforms and post-translational modifications on PIAS regulate its ability to self-associate (Figure 4–8) and form part of such complexes. PIAS interacting proteins that are also SUMOylation targets may be transient, occurring only at a specific point in the cell cycle or during development. 357 proteins were found to interact with PIAS in all conditions presented in this thesis

(Figure 5–7B) in the presence of NEM, indicating that these are constitutive binding proteins (not stage/condition specific).

PIAS interacts with proteins involved in: transcription, processing and transport of RNA and translation- indicating a regulatory role in all stages of gene expression. This has been identified qualitatively at the single protein level (Table 6) and in probabilistic models by gene ontology analysis (Table 5, Table 9, Table 10). This agrees with previous data that PIAS is essential for interphase chromosome organisation, chromosome structure, chromosome inheritance and gene expression[141].

In light of the recent data linking PIAS and Piwi-piRNA guided chromatin silencing[163] (an RNAi-guided mechanisms), it is of particular interest that members of the RNAi machinery were identified in PIAS IP-MS. Although, RNAi components have been classified as common mass spectrometry contaminants[343], this does not seem to be the case here. This is due to: large coverage of the pathway; the abundance of the factors following IP; and IgG controls with no detection of RNAi family proteins. Biologically, you would also expect RNA mediated transcriptional silencing complexes to be abundant, where it is important to protect from transposon activity.

Overlap of proteins identified in interactome studies is often limited. However, two independent mass spectrometry studies (discussed below), using HP1a and telomeric regions as bait, identified similar proteins as found in my mass spectrometry experiments. In my data, using both antibodies against endogenous PIAS and tagged versions of PIAS, there was a large overlap in equivalent conditions (Figure 4–22).

Published telomere-associated proteins include PIAS, SMT3, XNP and ADD1[322]. Many other proteins were also identified in this study and these overlap with PIAS interactors including HP1a, BEAF-32, Mi-2, and Caf1[322]. This overlap, combined with IF staining indicating localisation of PIAS to telomeres[143] and deficits in telomere clustering and telomere-lamina associations in *Su(var)2-10*^{1} *and Su(var)2-10*^{2} mutants demonstrates a role for PIAS in telomere function and nuclear organisation[141].

I confirmed an interaction between PIAS and HP1a (Figure 4–4) that had previously been found in HP1a IP-MS[299]. In mammals, SUMOylation of HP1α affects heterochromatin structure, increases HP1α chromatin binding affinity and enables targeting of HP1α to chromatin in the absence of H3K9me3. In a screen of HP1a interactors in *Drosophila* S2 cells, ADD1 was identified as the top hit of over 400 interactors, appearing in every condition tested[303]. SMT3 was also identified in 4/6 of their replicate studies[303]. XNP was not identified[303]. This study supports the presence of an HP1a-ADD1 and a HP1-SMT3 (SUMO) complex. It is also possible that ADD1 and SMT3 are part of a larger complex with HP1a, with SMT3 addition aided by PIAS as these proteins co-immunoprecipitate in the presence of NEM (Figure 7–5).

The transcription factors, without children (woc) and relative of woc (row) are consistently identified in this study to interact with PIAS, XNP and ADD1 (Figure 4–15, Figure 4–16, Figure 5–1, Figure 5–2, Figure 7–2 and Figure 7–6). Woc and row also co-immunoprecipitate with the boundary and insulator element BEAF-32[311]. Woc and row localise to developmental gene promotors with HP1c[311] and regulate telomere function[344,345].

8.4 The PIAS interactome changes during *Drosophila* development.

PIAS is present in high concentrations in the early embryo[142] where it localises to the DAPI dense regions of pre-blastoderm nuclei during heterochromatin establishment at nuclear cycles 10-14[144]. Kumar Hari observed PIAS localisation changing depending on the cell cycle stage[141] (Figure 2–4) and as having an inter-chromosomal thread-like localisation in S2 cells[141].

Different PIAS interactors were enriched from *Drosophila* embryos at different stages in development (Figure 5–4, Figure 5–5 and Figure 5–6), identifying unique PIAS interactors (Table 11, Table 13 and Table 15). During early development, PIAS interacts with pathways previously identified as regulators of heterochromatin formation including the Piwi-piRNA pathway and Eggless (Figure 5–1 and Table 11). These interactions are restricted to a specific developmental timepoint, suggesting changes in PIAS's interaction partners over time have an effect on PIAS's function.

Mass Spectroscopy data cannot be used to determine function of target proteins. It only provides information on the interactome of the protein or complexes it is involved in. From this, function can be inferred or suggested but must be followed up by phenotypic or functional analysis.

8.5 XNP interacts with PIAS

Previous experiments in the Heun lab suggested a close relationship between XNP and PIAS, as they co-localise in PIAS bodies in S2 cells. The lab therefore hypothesised PIAS could SUMOylate XNP because XNP can be SUMOylated[171]. We further hypothesised that failure to SUMOylate XNP due to loss of PIAS, could be responsible for the *Su(var)2-10*^{1}*/Su(var)2-10*^{2} phenotype, and of interest in alpha thalassemia x-linked mental retardation syndrome as a novel role for ATRX. However, interaction between XNP and PIAS was only observed at endogenous levels by immunoprecipitation of PIAS in the presence of NEM and not in reverse by immunoprecipitation of XNP. This is consistent with a role of SUMOylation of XNP by PIAS. In *Drosophila* tissues, PIAS co-localises with XNP at polytene inter-bands and at a repetitive region of the genome near heterochromatin on the X chromosome (at the TAGA repeat)[144]. However this interaction, compared with the many others identified in the PIAS interactome could be described as weak. A SUMO deficient mutant of XNP would be required to confirm whether the XNP-PIAS relationship was responsible for the *Su(var)2-10*^{1}*/Su(var)2-10*^{2} mutant phenotype.

8.6 ADD1 interacts with PIAS and is SUMOylated

IP-MS of PIAS interactors identified a novel interaction between PIAS and ADD1, which is independent of SUMOylation, indicating a non-SUMO dependent interaction (Figure 4–17). This was confirmed by IP-MS of ADD1, which identified PIAS. I also identified a SUMOylation site at K-146 on ADD1 (Figure 7–9).

8.7 ADD1 and XNP interact in the presence of SUMOylation

ADD1 is the *Drosophila* homologue of the N-terminal half of mammalian ATRX and XNP is the *Drosophila* homologue of the C-terminal half of mammalian ATRX[296]. In this study XNP and ADD1 only co-immunoprecipitate in the presence of NEM (Figure 7–5), suggesting that their interaction may be SUMO mediated or be part of a complex requiring SUMO. This allows a degree of protein regulation not present in

mammals, if we presume XNP and ADD1 fulfil the same function as mammalian ATRX when associated. PIAS, ADD1, XNP and HP1a co-immunoprecipitate when using any one of them as bait (Figure 7–4, Figure 4–4, Figure 4–13 and Figure 4–18A). This indicates that they might be part of a larger SUMO-dependent complex.

It is hypothesised that ADD1 localisation may maintain the border between heterochromatin and euchromatin domains[303] because during mitosis all HP1a interacting proteins except ADD1 are removed from chromatin[303] and during interphase ADD1 localises at the heterochromatin boundary and has weak, broad heterochromatin enrichment[303]. I have shown that ADD1 localises to DAPI-dense regions of pre-blastoderm nuclei where heterochromatin establishment takes place in the early *Drosophila* embryo (Figure 5–12). Although this does not prove that ADD1 marks areas for heterochromatin establishment during development, it indicates this is worthy of further study to determine if this is a novel and exciting role for ADD1 during development.

Similarly to ADD1, XNP localises to heterochromatin in narrow and focal subdomains in S2 cells[303]. However, unlike ADD1, XNP is removed during mitosis. My hypothesis is that ADD1 acts as the seed to recruit HP1 and XNP to sites of heterochromatin establishment, which is possibly aided by PIAS mediated SUMOylation. Therefore a PIAS, ADD1, XNP, HP1a interaction will only be present at specific points in development or in the cell cycle.

8.8 Speculative Models

In *Su(var)2-10*^{1}*/Su(var)2-10*^{2} trans-heterozygous mutants, interphase chromatin is decondensed[141]. My data indicates a number of novel interactions between PIAS and proteins involved in chromatin organisation, that could be required for establishing and/or maintaining properly folded chromatin. PIAS proteins potentially bind the nuclear lamina/matrix through their SAP domain[147]. Once at chromatin, PIAS may co-ordinate SUMOylation within a given chromatin domain either specifically or through a "SUMO spray" mechanism. Through this, PIAS may be essential to SUMOylate chromatin factors, ensuring their recruitment to form repressive chromatin complexes or to regulate transcription through SUMOylation of transcription factors[346]. In the presence of NEM, an increased number of protein-protein interactions were identified, indicating that SUMO promotes interactions between proteins. This may be particularly important during transition phases, such as during development, differentiation and diseases arising from defects in gene silencing, where chromatin must be restructured.

The first speculative model (Figure 9-1) is adapted from multiple sources and data gathered in this thesis. The original published figures copied without amendment are in Appendix 6. My model places PIAS in the context of its role in interphase heterochromatin organisation. A reduction in PIAS causes chromatin de-condensation in interphase nuclei[141]. PIAS localisation is similar to BEAF-32 (a boundary associated protein) and is associated with bordering regions of active transcription[57]. This indicates it could be found at the base of loops where transcription takes place. My IP-MS data identified many protein interactors of PIAS, and consistently identified proteins with roles related to RNA Pol II transcription. A long standing hypothesis of Peter Cook is that proteins involved in similar processes are clustered in "factories" to ensure efficient interaction[347-349].

In my model I have placed PIAS at the base of chromatin loops close to RNA Pol II transcription factories, where it could function in several related pathways or play dual roles as both an activator or repressor based on chromatin context. My data demonstrates that PIAS interacts with a larger number of proteins in the presence of NEM. This indicates SUMO reliant (SUMO-SIM) interactions (sometimes referred to as “glue”), known to bind proteins into large networks (see Appendix model 8)[350]. PIAS mediated SUMOylation presents a theoretically efficient method to “glue” proteins in SUMO-SIM networks[350] (at the base of loops) thereby causing locally increased factor concentrations required for chromatin establishment and formation. In short, localisation of PIAS at the base of loops may limit PIAS’ SUMO conjugation capacity to target chromatin proteins within its immediate environment, “gluing” them at sites of active transcription, generating a high local concentration required for establishing chromatin domains, determining which areas of the genome are actively transcribed or silenced.

An outstanding question in the SUMOylation field is why are only a small sub-fraction of proteins SUMOylated. This model could explain why only a subset of proteins are SUMOylated, as only those in PIAS’s immediate vicinity and required at the time become SUMOylated.

This page is intentionally left blank

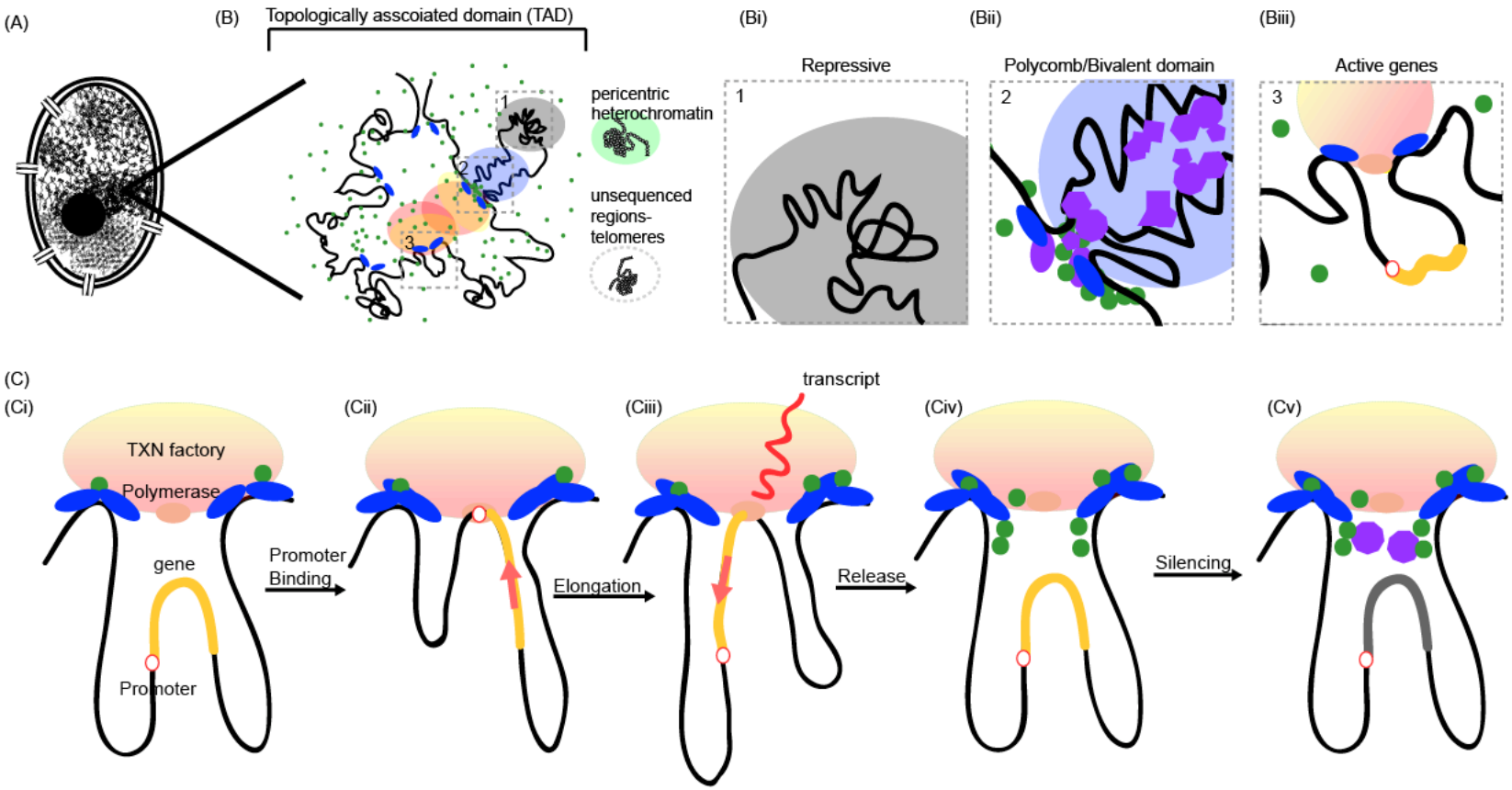

(A)
(B)
Topologically asscoiated domain (TAD)
pericentric heterochromatin
unsequenced regions- telomeres
(Bi)
Repressive
(Bii)
Polycomb/Bivalent domain
(Biii)
Active genes
(C)
(Ci)
TXN factory
Polymerase
gene
Promoter
(Cii)
Promoter Binding
(Ciii)
transcript
Elongation
(Civ)
Release
(Cv)
Silencing

Figure 8–1 Regulation of Insulator Function by PIAS-mediated SUMOylation
(A) Chromatin is organised within domains in the nucleus. **(B)** A topologically associated domain made up of chromatin (black) in open and closed formats with interspersed SUMO (green), and PIAS (blue) bound to euchromatin, particularly at the base of loops. Areas of functionally differential chromatin are illustrated by coloured ovals; HP1 (GREEN), Active (YELLOW/RED), Polycomb (BLUE) or Repressive (BLACK). **(Bi)** Inset 1 from B showing a repressive domain **(Bii)** Inset 2 from B showing a Polycomb/Bivalent loop in transition to repressive, with chromatin (black) forming a loop with PIAS (blue) at the base and a high local concentration of SUMO (green) which recruits and maintains a high local concentration of chromatin factors (purple) required for remodelling from Polycomb/Bivalent to repressive chromatin. **(Biii)** Inset 3 from B showing an active gene with chromatin (black) forming a loop with PIAS (blue) at the base with a low local concentration of SUMO (green), an open gene (yellow) with promotor (red circle) is available for transcription. **(C)** Biii showing gene transcription followed by silencing. **(Ci)** PIAS sits at the base of active chromatin loops and helps recruit RNA polymerase to a transcription factory (TXN factory). The gene of interest and promoter sit in open chromatin within the loop. **(Cii)** The gene promoter is recruited to the RNA polymerase complex. **(Ciii)** The gene is pulled through the RNA polymerase complex, producing transcript. **(Civ)** The transcript is released and chromatin binding proteins (histones and insulators proteins) are SUMOylated, parting the loop, causing a more open chromatin conformation. **(Cv)** Proteins involved in forming repressive complexes can be recruited by SUMO, the promoter is no longer accessible and transcription from the gene is prevented through isolation of the loop.

Though the model above is based upon published literature there are a number of outstanding questions:

- how chromatin factors are initially recruited to chromatin to seed heterochromatin formation in *Drosophila*
- how decisions are made as to which factors to recruit
- how heterochromatin is nucleated and maintained.

Although my interactome data only implies relationships without functional confirmation, from it and with support from the literature I can form a working hypothesis for heterochromatin formation in the early *Drosophila* embryo.

Models for heterochromatin establishment are discussed in sections 1.11 and 2.3.13. Heterochromatin is defined at the molecular level by the presence of the histone modification H3K9me3 and HP1. HP1 proteins can be recruited to form heterochromatin by multiple mechanisms and HP1 binding is enhanced in a positive feedback loop to form heterochromatin domains.

In *S. pombe* heterochromatin is established through RNA interference (RNAi) based mechanisms, using transcripts transcribed from heterochromatic regions. These transcripts target histone methyltransferases to deposit methyl marks on histone 3 lysine 9 (Figure 1-2)[351] to form heterochromatin. In *Drosophila* PIAS and XNP are required for transcription of pericentric repeats[144] that could act in RNAi-based mechanisms to maintain heterochromatin. Furthermore, PIAS interacts with Piwi, a Piwi-piRNA component known to promote heterochromatin assembly (Figure 1–3). Piwi-piRNA were identified in PIAS IP-MS only at the earliest embryo timepoints (Figure 5–1) and RNAi proteins were identified in all PIAS IP-MS conditions.

PIAS IP-MS in timepoint 1 embryos (nuclear cycles 7-10) also identified the histone methyltransferase Eggless, which Seller *et al.,* 2019[123] propose to seed the formation of heterochromatin by depositing methyl marks at H3K9. The mechanism of Eggless recruitment to chromatin is currently unknown. Seller *et al.,* 2019[123] propose interaction of Eggless with small RNAs could guide Eggless to chromatin. Given the novel interaction presented in this thesis between PIAS and Eggless (Table 11), it is possible that PIAS-mediated SUMOylation of Eggless could enhance its association with chromatin. This would require further confirmation. In support of this hypothesis, similar mechanisms have been identified. A SUMO dependent mechanism for *de novo* recruitment of HP1α to heterochromatin in mouse ESCs was identified by Maison *et al.,* 2011[244], which also requires pericentric long non-coding RNAs that correspond to major satellite repeats. Subsequently, the same researchers found the histone methyltransferase Su39h1 can exhibit novel E3 SUMO ligase activity to provide this SUMOylation[247].

In *Drosophila*, although HP1a binding partners have been identified, none are E3 SUMO ligases[299]. An argument for PIAS's involvement in heterochromatin formation in *Drosophila* is that it interacts with proteins required for heterochromatin establishment at the timepoints when SUMOylated versions of the protein are required for recruitment to or function of at chromatin. Eggless interacts with HP1a[299] and ADD1[299]. ADD1, which can directly bind DNA, may provide an alternative means of HP1a recruitment and act to nucleate heterochromatin formation in a mechanism analogous to mammalian ATRX, which has been shown to interact with HP1α[352]. A combination of redundant mechanisms may act to initiate formation of and stabilisation of heterochromatin domains.

Below I outline a hypothetical model for how ADD1, PIAS, XNP and H3.3 are involved in incorporation of H3.3 at chromatin (Figure 8–2).

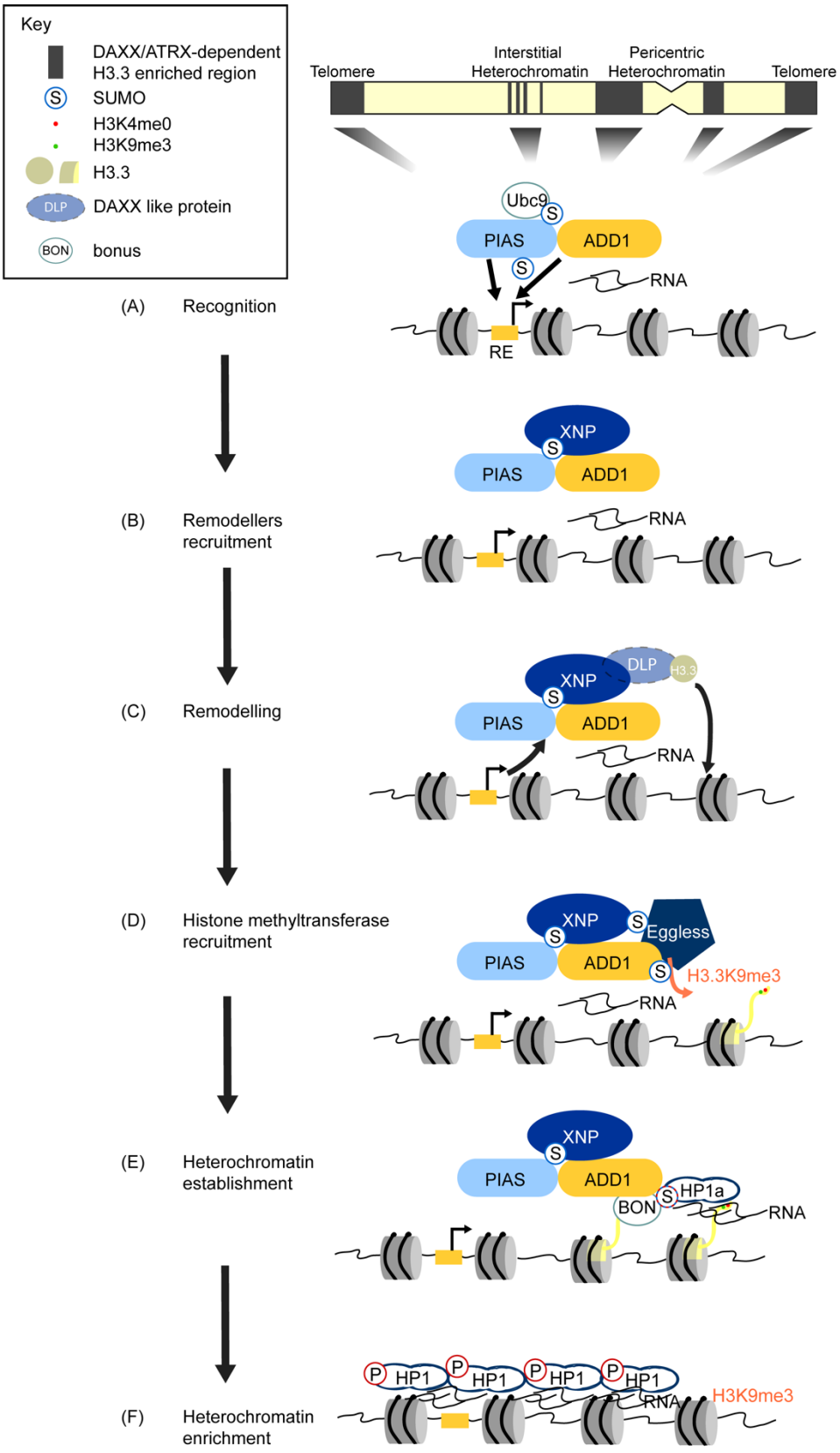

Key
DAXX/ATRX-dependent H3.3 enriched region
SUMO
H3K4me0
H3K9me3
H3.3
DLP
DAXX like protein
BON
bonus
Telomere
Interstitial Heterochromatin
Pericentric Heterochromatin
Telomere
Ubc9
PIAS
ADD1
RNA
RE
(A) Recognition
XNP
(B) Remodellers recruitment
DLP
H3.3
(C) Remodelling
Eggless
H3.3K9me3
(D) Histone methyltransferase recruitment
HP1a
(E) Heterochromatin establishment
HP1
(F) Heterochromatin enrichment

Figure 8–2 Hypothetical working model of PIAS, ADD1 and XNP at the equivalent of DAXX/ATRX-dependent H3.3 enriched regions of heterochromatin in *Drosophila*.
(A) PIAS localises to recognition elements in DNA (RE). ADD1 can directly bind DNA and interact with PIAS independently of its SUMOylation status (Table 6). ADD1 binding is reinforced by binding of the ADD domain to a combination of histone tail modifications (as for human ATRX[258]). Both PIAS and ADD1 interact with components of the TFIID complex (Figure 4–15 and Figure 7–6) and are likely to be involved in mediating transcription close to their binding sites in the genome. For example, ADD1 prevents expression of TART retroelements in the telomeric HTT array[295]. PIAS is able to autoSUMOylate[144] and SUMOylate factors that come into proximity with it in a "SUMO spray" type mechanism[225]. In collaboration with this, RNA may act as scaffolding or target sequences in RNAi pathways to both structurally and specifically direct chromatin remodelling. Together, RNA and SUMOylation based mechanisms could help provide an open chromatin configuration, enabling recruitment of chromatin remodellers, as is the case for SUMOylation of histone 4[234,235]. **(B)** Proteins that are SUMOylated or contain SIM domains can form multivalent interactions with each other and are retained at chromatin. XNP can be SUMOylated[171] and when SUMOylated binds to ADD1 (Figure 7–6). **(C)** XNP is involved in the deposition of histone H3.3 in nucleosomes through interaction with DAXX-like protein (DLP)[115] at pericentric and telomeric repeats to maintain an open conformation of DNA[353]. **(D)** The histone methyltransferase Eggless (SETBD1 homolog) is recruited and tri-methylates H3K9. Eggless can be modified by SUMO[354], interacts with PIAS (Table 11) and ADD1[299]. Koch *et al.,* 2009 propose that SUMOylated Eggless can recruit chromatin regulators and form repressive chromatin complexes[354]. **(E)** H3K9me3 and ADD1 recruit HP1a independent of Eggless[355] and ADD1 is required for HP1a recruitment to telomeres[295] (in mammals there is cross-talk between DNA methylation and histone methylation, *de novo* DNA methyltransferases can interact with histone methyltransferases[356] to cooperate for the repression of gene expression[357]). BON could be responsible for HP1a hinge domain phosphorylation in *Drosophila*, as mammalian homologs perform this function[249]. **(F)** HP1a's hinge domain is important for its DNA binding properties in *Drosophila*[358]. In mammals, phosphorylation of this region alters HP1α DNA binding stability and interaction partners [359]. The mouse HP1α hinge domain has also been shown to be SUMOylated[244]. It is possible that SUMO-phosphorylation cross-talk happens in the hinge region of HP1 in a cell cycle or cell stage dependent manner[352] to strengthen and reinforce HP1 domains in feedback loops[124]. HP1 forms and defines heterochromatin domains.

8.9 Future experiments

8.9.1 To understand the global role of PIAS at chromatin

In principle, with tagged versions of PIAS and endogenous antibodies that can immunoprecipitate PIAS from different cell types and tissues, the following experiments would provide support for a global role of PIAS at chromatin.

8.9.1.1 Mapping PIAS genomic binding sites

PIAS DamID, which maps protein binding sites in the genome, has been performed in the embryonic *Drosophila* Kc167 cell line[57]. This revealed that PIAS binds to euchromatic regions[57]. PIAS binding profiles could be explored to identify genomic features of PIAS binding sites. If this identified G4 structures and transcription start sites, it would along with my interactome data (which identified Tafs and transcription factors), builds a case of PIAS mediated transcription regulation in *Drosophila*. PIAS DamID or ChIP-seq could be performed from *Drosophila* embryos to identify changes in PIAS binding sites through development. This localisation data could aid study of the function of PIAS in establishing chromatin domains (e.g. at heterochromatin) before PIAS dissipates and binds to euchromatin (as observed in tissue culture cell lines).

8.9.1.2 Identifying sites of PIAS binding on RNA

In light of the data identifying interactions between PIAS and RNA binding proteins (Figure 5–11), it would be interesting to investigate whether PIAS associates directly with RNA. To identify how PIAS is involved in RNA processing, PIAS cross-linking and analysis of cDNAs (CRAC) or RNA immunoprecipitation sequencing (RIPseq) could be performed. This would allow identification of the RNA bound component of PIAS complexes and identify if there is a consensus sequence on mRNAs where PIAS binds. This would allow identification and classification of the RNAs that bind PIAS and importantly it would allow a comparison to the RNAs identified by HP1a CRAC[299] to determine whether there are specific RNAs acting at heterochromatin regions of the genome.

8.9.1.3 Identifying dynamic changes in PIAS localisation during development

Antibodies used in immunofluorescence on fixed embryos can be used as a starting point to identify the timing of localisation, nucleation sites and how heterochromatin domains expand by comparison to HP1a domain formations. However, given the current perspective of heterochromatin organisation into domains based on the biophysical properties of chromatin factors and associated proteins, only by looking in *Drosophila* embryos during development over a time course can we being to understand the functional importance of PIAS proteins in organisation of the genome. Phase separation of heterochromatin plays a critical role in nuclear organisation through dynamic changes and building local concentrations of chromatin factors[124]. PIAS could be fluorescently tagged in the *Drosophila* genome to allow live-imaging of *Drosophila* embryos to track and dissect the role of PIAS in these processes. CRISPR/Cas9 targeting in S2 cells provided a proof-of-principle that PIAS can be tagged at the endogenous *Su(var)2-10* locus. Using the same targeting strategy to generate a *Drosophila* line would allow study of PIAS localisation using high-resolution light microscopy without overexpression of PIAS.

8.9.2 To understand the role of PIAS complexes

Investigations of PIAS complexes could be pursued. PIAS elutes from gel filtration columns over a wide volume, indicating its involvement in complexes of various sizes. IP-MS from these different PIAS fractions could inform on the composition of PIAS subcomplexes.

8.9.3 To study PIAS function by knockout/knockdown

PIAS KO is lethal (Heun lab unpublished). Therefore, transient knockdown studies to assay the phenotype of PIAS reduction are required. This would probably require a rapid, acute depletion strategy. This is an ongoing effort in the Heun lab.

8.9.4 To study PIAS function in genome organisation

Hi-C experiments on a PIAS inducible knockout or a PIAS E3 ligase deficient mutant would allow investigation of the role of PIAS in chromatin structural organisation to aid understanding in the context of chromatin loop formation by PIAS and PIAS mediated SUMOylation. This would be helpful towards determining a role for PIAS and PIAS's enzymatic activity in establishing chromatin domains.

Domain deletion mapping could be used to assess the importance of PIAS functional domains and map sites of interaction with HP1a and ADD1 within PIAS.

8.9.5 To understand the role of SUMOylation of XNP and ADD1

Deciphering the mechanism of action of XNP-SUMOylation is important for understanding SUMO regulation of XNP function. XNP and ADD1 may confer *Su(var)* phenotypes in the ln(1)w^{m4} PEV assay through their interaction with components of the SUMOylation machinery. Mass Spectrometry identification of SUMOylation sites on ADD1 and XNP allows future targeting of sites to generate ADD1 and XNP SUMOylation mutants. This would be required to address the function of the SUMO-mediated ADD1 and XNP interaction at both the protein and chromatin levels.

8.9.6 To identify an ADD1-XNP complex

Study of ADD1, XNP and SMT3 in *Drosophila melanogaster* has both advantages and disadvantages. The main advantage of *Drosophila*, as opposed to mammalian cell culture, is the simpler SUMO machinery and fewer redundancies in proteins and mechanisms so multiple knockouts are not required. However, in *Drosophila* the functions of ADD1 and XNP as opposed to the single ATRX must be addressed separately. Overall *Drosophila* allows more precise manipulation but has important molecular differences to human cells and it is not clear how far these extend at a functional level. Using genome editing, ADD1 and XNP could be fused or mammalian ATRX could be expressed in *Drosophila* to compare the function of ADD1 and XNP to ATRX and explore the differences in regulation that may have arisen as a result of the gene split.

It is possible that ADD1 and XNP act in distinct but overlapping pathways in *Drosophila* compared to vertebrate ATRX proteins. It would be interesting to explore whether ADD1 is responsible for “bookmarking” heterochromatin, possibly through mechanisms analogous to *de novo* DNMTs.

8.10 Thesis Summary

I began with four key aims:

1. to develop tools for identifying PIAS, ADD1 and XNP interactomes at endogenous protein levels
2. to understand whether the previously identified PIAS interaction with XNP occurs under physiological conditions
3. to characterise the interaction between ADD1 and XNP with PIAS
4. to identify other PIAS interactors that may have an effect at heterochromatin.

Tools for investigating the PIAS, ADD1 and XNP interactomes at endogenous levels were successfully developed, these allowed me to confirm the previously identified PIAS-XNP interaction and identify an interaction between PIAS and ADD1. It was also found that ADD1 and XNP co-immunoprecipitate only in the presence of NEM, suggesting that their interaction is mediated by SUMOylation. Finally the interactomes of PIAS, ADD1 and XNP were identified along with a large number of novel interactors.

The major findings of this thesis are:

- XNP interacts with PIAS at endogenous expression levels but only in the presence of SUMOylation.
- ADD1 is SUMOylated and localises to areas of heterochromatinization in the early embryo.
- GO terms analysis indicates large numbers of PIAS interactors involved in DNA, RNA binding and transcription. This matches previous data indicating SUMOylation is essential to all stages of gene regulation.
- PIAS is a potential novel regulator of heterochromatin establishment through interaction with the developmental/germline specific piRNA pathway.
- PIAS interacts with Eggless. The mechanism of Eggless targeting to heterochromatin is currently unknown. SUMOylation is known to target proteins to specific localities and is an exciting potential mechanism for the function of Eggless.

8.11 General conclusion

Establishment and maintenance of heterochromatin is vital to cellular function and cellular viability. Dysregulation of chromatin factors, such as ATRX, results in diseases such as Alpha-Thalassemia X-linked mental retardation syndrome and cancers. SUMOylation of heterochromatin-associated proteins has been identified as an initiation step in seeding heterochromatin formation, potentially by aiding localisation to heterochromatin. In this thesis a number of novel interactors of the SUMO E3 ligase PIAS have been identified, including ADD1 and XNP (Human ATRX homologs) and other chromatin factors. I suggest that the interactome of PIAS indicates that it may be a key factor to control and recruit proteins involved in multiple mechanisms of chromatin silencing and remodelling. These mechanisms are involved in genome packaging, which regulates gene expression, and is often disrupted in cancer cells.

Appendices

Appendix 1: Mass Spectrometry Optimisation and Quality Control (QC)

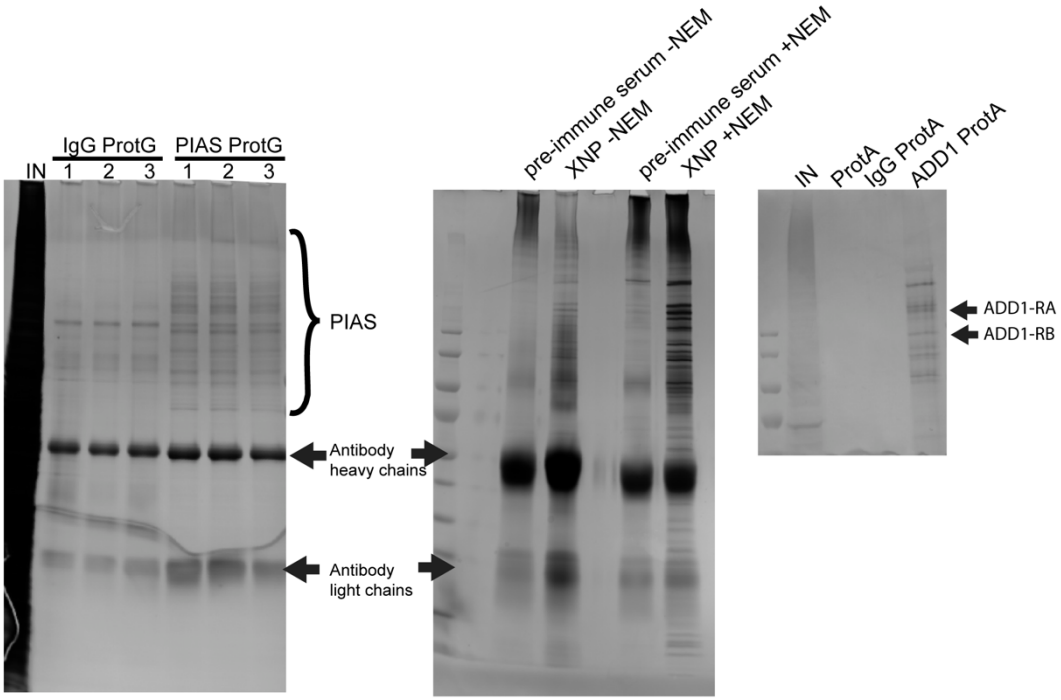
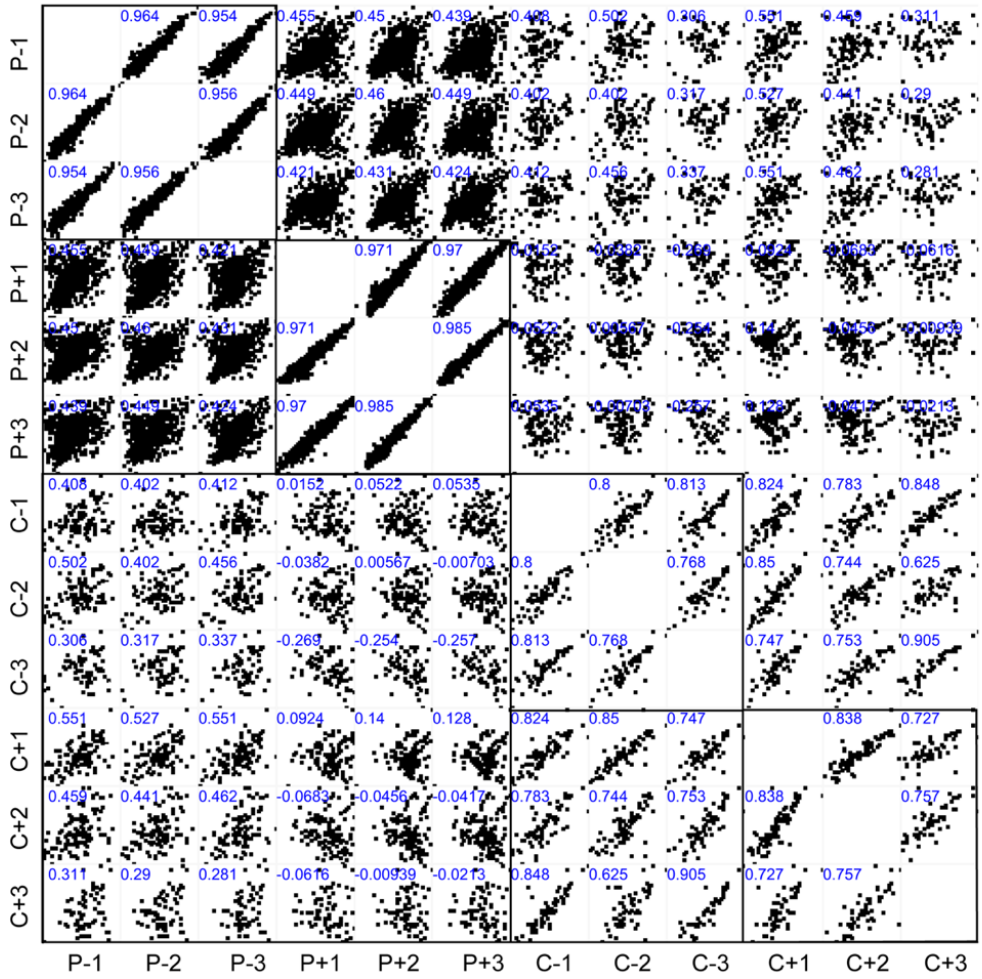


Figure 8–3 4-20% SDS-PAGE silver stained gels for PIAS, XNP and ADD1 IP

IgG and PIAS immunoprecipitations in triplicate from WT S2 cells for mass spectrometry carried out in the absence of N-ethylmaleimide (NEM). XNP IP using pre-immune serum or XNP bleed serum in the absence and presence of NEM. ADD1 immunoprecipitates in the ADD1 ProtA lane but is absent from other lanes. IP samples used for Mass Spectrometry of the PIAS, XNP and ADD1-associated complexes. Bands at 50 and 20 kDa correspond to the detached and denatured heavy and light chains of the antibodies, which were present in excess.



	P-1	P-2	P-3	P+1	P+2	P+3	C-1	C-2	C-3	C+1	C+2	C+3
P-1		0.964	0.954	0.455	0.450	0.439	0.408	0.502	0.306	0.551	0.459	0.311
P-2	0.964		0.956	0.449	0.460	0.449	0.402	0.402	0.317	0.527	0.441	0.290
P-3	0.954	0.956		0.421	0.431	0.424	0.412	0.456	0.337	0.551	0.462	0.281
P+1	0.455	0.449	0.421		0.971	0.970	0.015	-0.038	-0.269	0.092	-0.068	-0.061
P+2	0.450	0.460	0.431	0.971		0.985	0.052	0.006	-0.254	0.140	-0.046	-0.009
P+3	0.439	0.449	0.424	0.970	0.985		0.054	-0.007	-0.257	0.128	-0.042	-0.021
C-1	0.408	0.402	0.412	0.015	0.052	0.054		0.800	0.813	0.824	0.783	0.848
C-2	0.502	0.402	0.456	-0.038	0.005	-0.007	0.800		0.768	0.850	0.744	0.625
C-3	0.306	0.317	0.337	-0.269	-0.254	-0.257	0.813	0.768		0.747	0.753	0.905
C+1	0.551	0.527	0.551	0.0924	0.140	0.128	0.824	0.850	0.747		0.838	0.727
C+2	0.459	0.441	0.462	-0.068	-0.045	-0.041	0.783	0.744	0.753	0.838		0.757
C+3	0.311	0.290	0.281	-0.062	-0.009	-0.021	0.848	0.625	0.905	0.727	0.757	

Figure 8–4 Multi-scatter plot of pairwise comparisons of protein abundance (LFQ) detected in each PIAS IP-MS sample from S2 cells

PIAS (P) or IgG (C) immunoprecipitation from wild-type S2 cell nuclear extracts followed by mass spectrometry were performed in the absence (-) or presence (+) of NEM in triplicate. Each IP-MS sample is represented along each axis. Each dot represents one protein plotted by its $\log_2$ transformed label-free quantification (LFQ) value on an axis spanning values of 20 to 30. Larger boxes in black along the diagonal show groupings of replicates paired against each other. Pearson correlation values are shown in the top left-hand corner of each box and below.

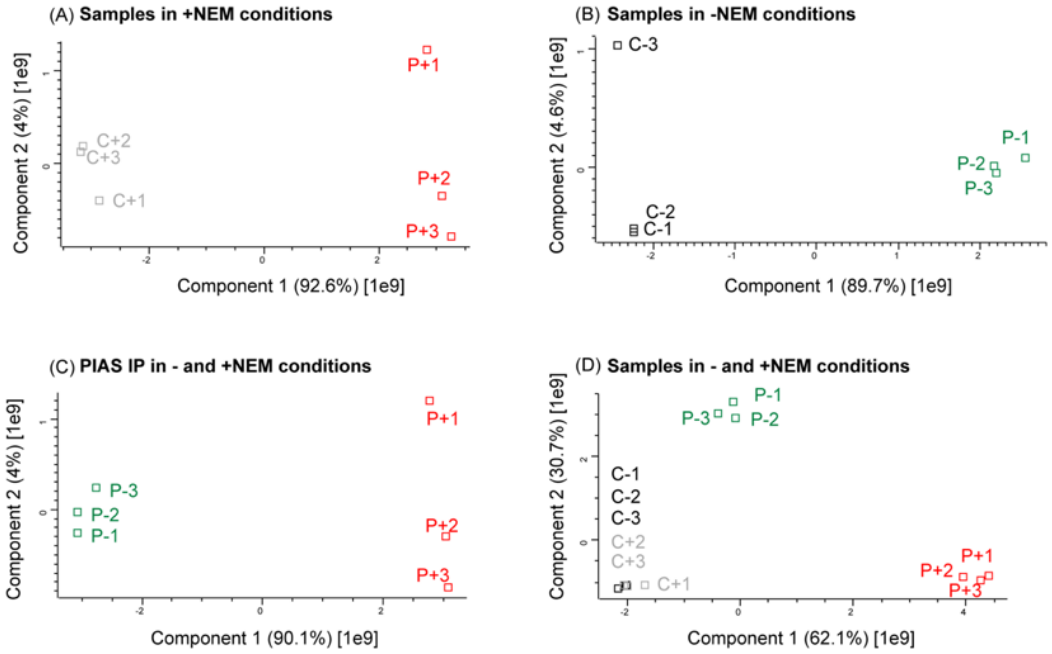


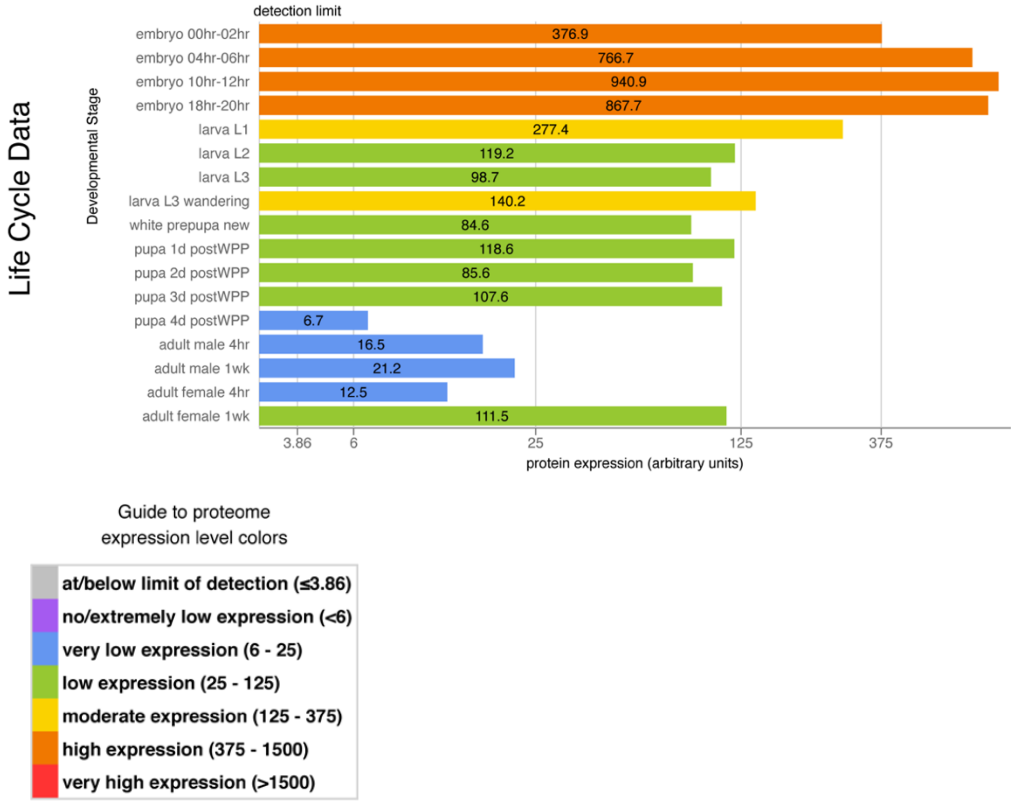
Figure 8–5 Principle Components analysis of PIAS IP-MS samples from S2 cells

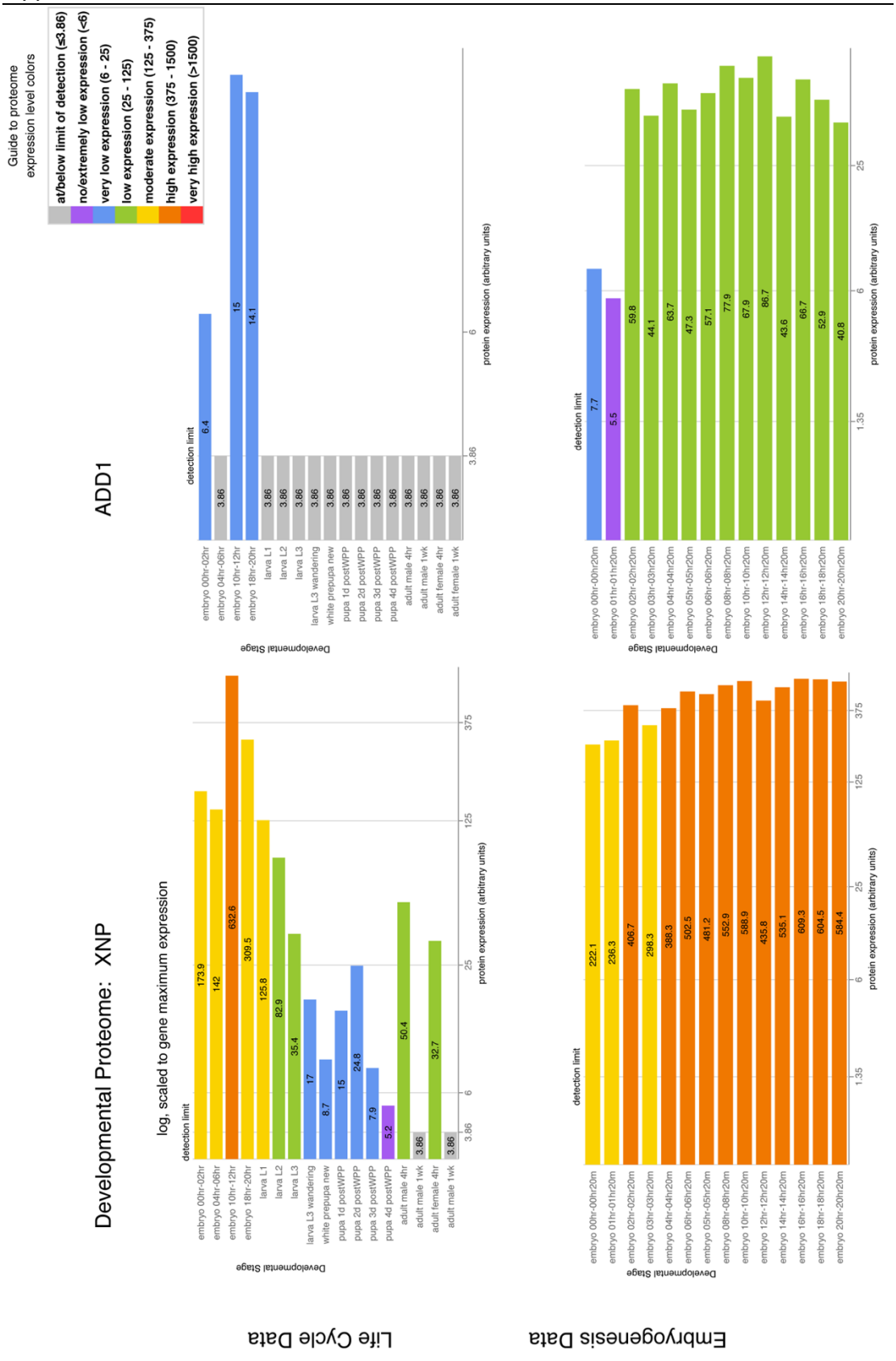
(A and B) showing separation of samples from controls **(C)** samples with or without NEM and **(D)** samples with or without NEM separating from controls **Key:** PIAS IPs in NEM conditions (P+), PIAS IPs with no NEM (P-), IgG IPs in NEM conditions (C+), IgG IPs with no NEM (C-).

Appendix 2: PIAS, ADD1 and XNP Expression Levels and Limits of Detection

Figure 8–6 Protein expression levels of PIAS, XNP and ADD1

from a quantitative proteomics study of *Drosophila* embryos at different stages in development[142] downloaded from FlyBase 04-04-2019. Protein expression levels are in arbitrary units. Expression level classification is colour coded in the key.




Guide to proteome expression level colors
at/below limit of detection (≤3.86)
no/extremely low expression (<6)
very low expression (6 - 25)
low expression (25 - 125)
moderate expression (125 - 375)
high expression (375 - 1500)
very high expression (>1500)
ADD1
detection limit
6.4
3.86
15
14.1
3.86
Developmental Stage
protein expression (arbitrary units)
Developmental Proteome: XNP
log, scaled to gene maximum expression
173.9
142
632.6
309.5
125.8
82.9
35.4
17
8.7
15
24.8
7.9
5.2
50.4
32.7
Life Cycle Data
Embryogenesis Data
7.7
5.5
59.8
44.1
63.7
47.3
57.1
77.9
67.9
86.7
43.6
66.7
52.9
40.8
222.1
236.3
406.7
298.3
388.3
502.5
481.2
552.9
588.9
435.8
535.1
609.3
604.5
584.4

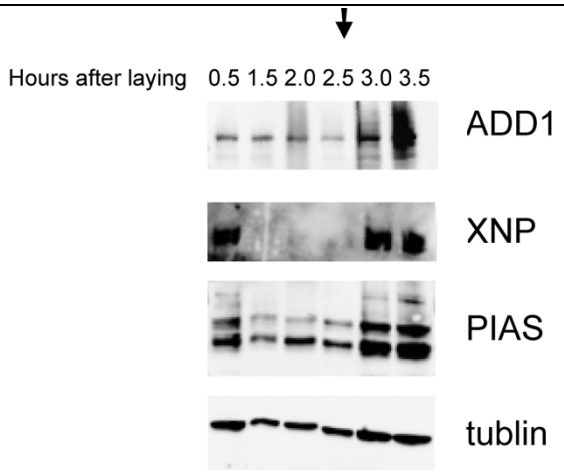


Figure 8–7 Total protein extract from 50 *Drosophila* embryos
Western blot probed with antibodies recognising ADD1, XNP, PIAS and tubulin. Embryos were prepared by manual puncture and boiling in 2X LB, at the specified time points.

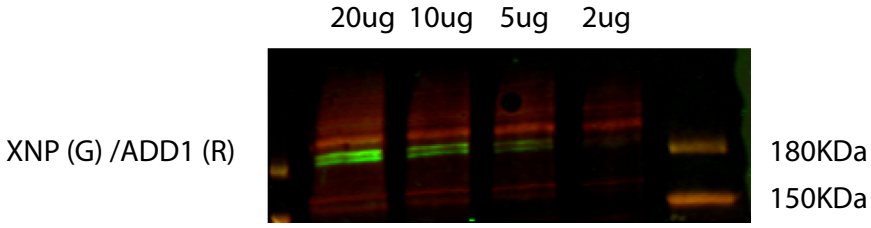


Figure 8–8 Detection limit of XNP and ADD1 in lysate from *Drosophila* embryos
Nuclear extract from wild-type S2 cells serially diluted and run on 6% SDS-PAGE for western blot probed with anti-ADD1 rabbit bleed 3 serum (in red) and anti-ATRX rat serum (in green).

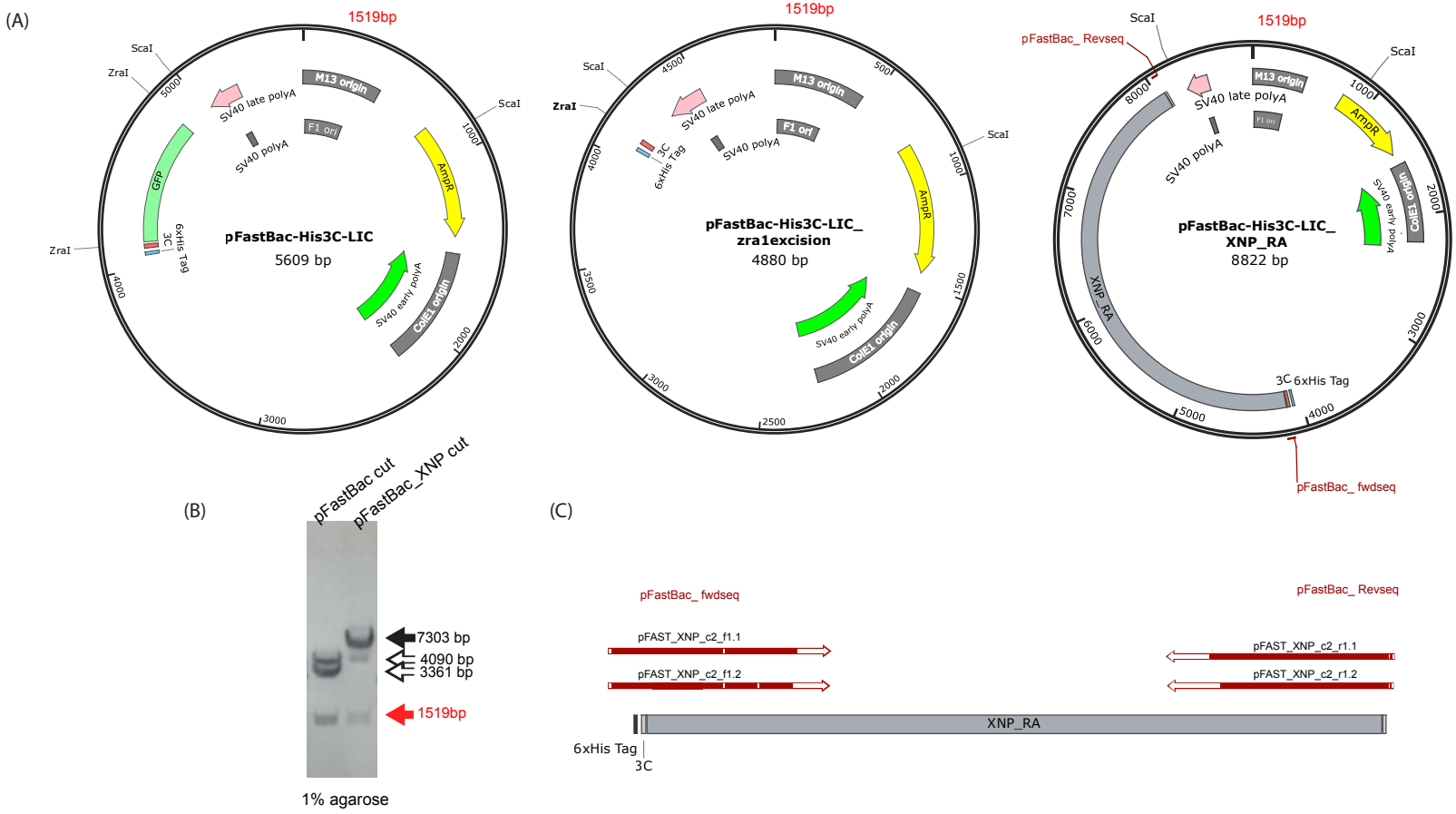
Appendix 3: Generation of Plasmids for ADD1 and XNP Protein Expression

Figure 8–9 ScaI restriction digest of pFastBac-HIS3C-LIC_XNP_RA to confirm plasmid construct

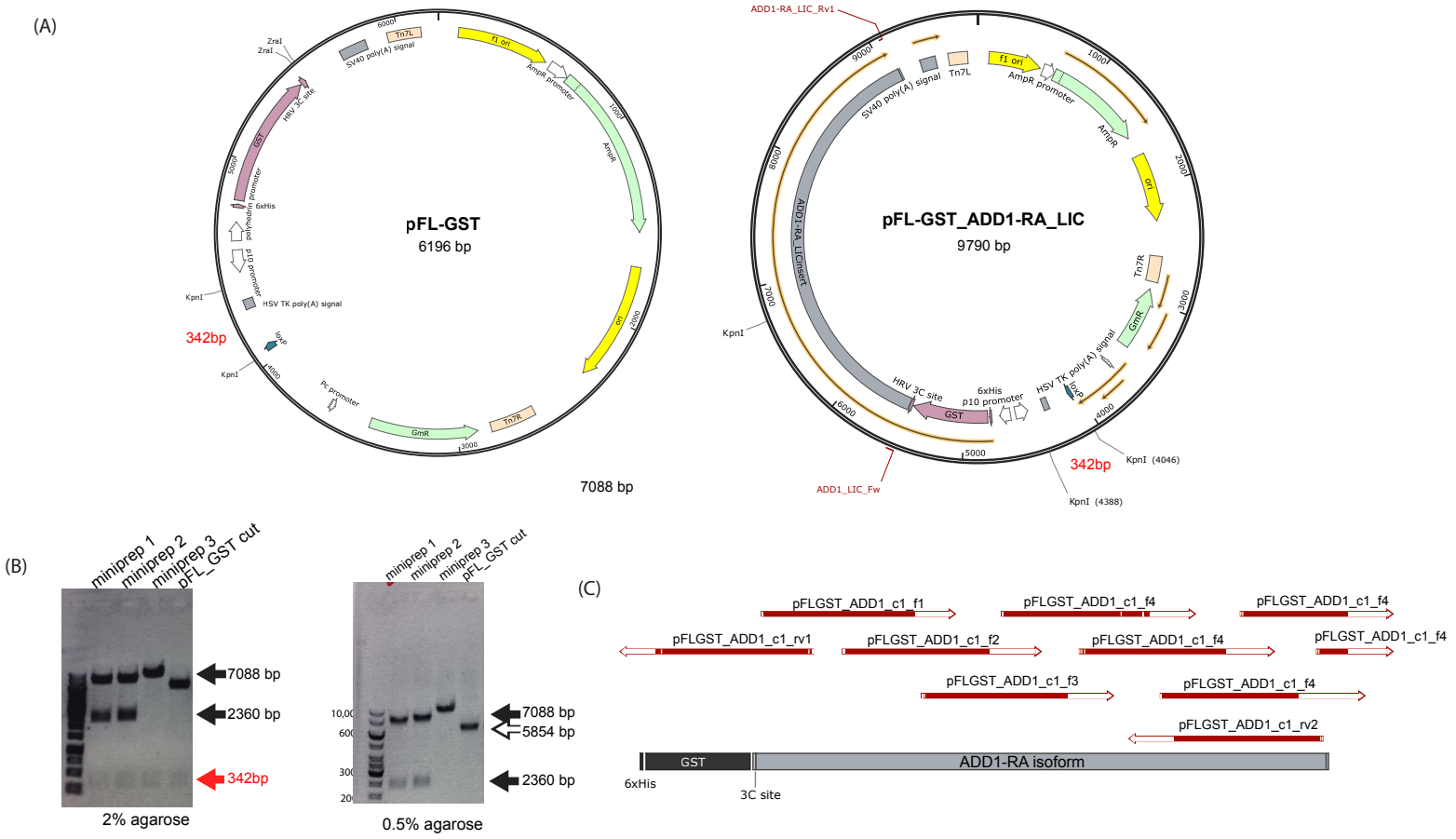
(A) Plasmid maps of pFastBac-HIS3C-LIC and pFastBac-HIS3C-LIC_XNP_RA marked with restriction sites used in this study. Two Zra1 sites are present in pFastBac-HIS3C-LIC which can be used to excise the GFP fragment for use in LIC cloning of other DNA fragments into this region of the vector. Two ScaI sites are present in pFastBac-HIS3C-LIC and pFastBac-HIS3C-LIC_XNP_RA. **(B)** Restriction digest of pFastBac-HIS3C-LIC results in two bands on agarose gel electrophoresis: 1519 and 3361 or 4090bp depending on whether the GFP fragments has been excised with zra1 prior to digestion with ScaI. Restriction digest of pFastBac-HIS3C-LIC_XNP_RA results in two bands on agarose gel electrophoresis: 1519 and 7303bp. The molecular weight size marker is not shown as this is part of a larger screening gel. **(C)** Sanger Sequencing reads aligned to the DNA sequence of XNP-RA confirming insertion of XNP-RA sequence into the pFastBac vector.

Figure 8–10 Kpn1 restriction digest of pFL-GST_ADD1-RA_LIC to confirm plasmid construct

(A) Plasmid maps of pFL-GST and pFL-GST_ADD1-RA_LIC marked with restriction sites used in this study. Two Kpn1 sites are present in pFL-GST and three Kpn1 sites are present in pFL-GST_ADD1-RA_LIC. **(B)** Restriction digest of pFL-GST_ADD1-RA_LIC results in three bands on agarose gel electrophoresis: 342, 2360 and 7088bp seen in minipreps 1 and 2. Restriction digest of pFL-GST results in two bands on agarose gel electrophoresis: 342 and 5854bp seen in the final lane. Half the restriction digest was run on a 2% agarose gel and the other half on a 0.5% agarose gel to visualise the resolution of the bands against the molecular weight size marker (Thermo Fisher Scientific™ GeneRuler™ 1kb DNA Ladder). **(C)** Sanger Sequencing reads aligned to the DNA sequence of ADD1-RA confirming insertion of the correct sequence of ADD1-RA sequence into the pFL vector.


(A)
1519bp
ScaI
ZraI
M13 origin
SV40 late polyA
F1 ori
SV40 polyA
GFP
3C
6xHis Tag
AmpR
SV40 early polyA
ColE1 origin
pFastBac-His3C-LIC
5609 bp
pFastBac-His3C-LIC_
zra1excision
4880 bp
pFastBac-His3C-LIC_
XNP_RA
8822 bp
pFastBac_ Revseq
pFastBac_ fwdseq
(B)
pFastBac cut
pFastBac_XNP cut
7303 bp
4090 bp
3361 bp
1519bp
1% agarose
(C)
pFastBac_ fwdseq
pFastBac_ Revseq
pFAST_XNP_c2_f1.1
pFAST_XNP_c2_f1.2
pFAST_XNP_c2_r1.1
pFAST_XNP_c2_r1.2
XNP_RA
6xHis Tag
3C

Appendices


(A)
pFL-GST
6196 bp
342bp
7088 bp
pFL-GST_ADD1-RA_LIC
9790 bp
ADD1-RA_LIC_Rv1
ADD1_LIC_Fw
342bp
KpnI (4046)
KpnI (4388)
(B)
miniprep 1
miniprep 2
miniprep 3
pFL_GST cut
7088 bp
2360 bp
342bp
2% agarose
7088 bp
5854 bp
2360 bp
0.5% agarose
(C)
pFLGST_ADD1_c1_f1
pFLGST_ADD1_c1_f4
pFLGST_ADD1_c1_rv1
pFLGST_ADD1_c1_f2
pFLGST_ADD1_c1_f3
pFLGST_ADD1_c1_rv2
GST
ADD1-RA isoform
6xHis
3C site

Appendix 4: Generation of an HA-ADD1 S2 Cell Line

To provide an alternative approach for isolating ADD1 complexes for validation of ADD1 interactors, ADD1 HA-tagged S2 cell lines were generated, screened and validated. The HA tag was inserted in ADD1 exon 1 (Figure 8–11A) to tag all three isoforms of ADD1. A schematic of the CRISPR/Cas9 strategy used for insertion of the tag is shown in Figure 8–11B and the homology directed repair template design in Figure 8–11C.

Of the screen of S2 cell clones for insertion of the tag, clone 1 has the HA-tag integrated in all four copies of the genome at the N-terminus (Figure 8–11D). Clone 2 contains at least one tagged allele and 3 contains no tagged alleles. HA-ADD1 S2 cell clones were also screened for HA expression by detection of the HA epitope using HA antibodies able to be detected by immunofluorescence (Figure 8–11E). This revealed that clone 2 is a mixed population as only some cells express HA and that clone 3, in agreement with the PCR, expresses no HA.

Western blot probed with an anti-HA antibody recognised protein bands of the expected size in protein lysate generated from HA-ADD1 cell line 1 that were not observed in protein extracts from wild-type S2 cells (Figure 8–11F). Therefore clone 1 was used for further study. Protein extract from HA-ADD1 cell line 1 was subject to immunoprecipitation with anti-HA-resin, but with high background (Figure 8–11G) in NEM conditions, shown by probing the bottom half of the membrane with *Drosophila* anti-PIAS sheep antibody, as PIAS should co-immunoprecipitate with ADD1 in the HA-ADD1 cell line 1. PIAS should not be observed in the wild-type (WT) lane, as this represents non-specific binding to the anti-HA resin as no HA species are present in the WT cell line.

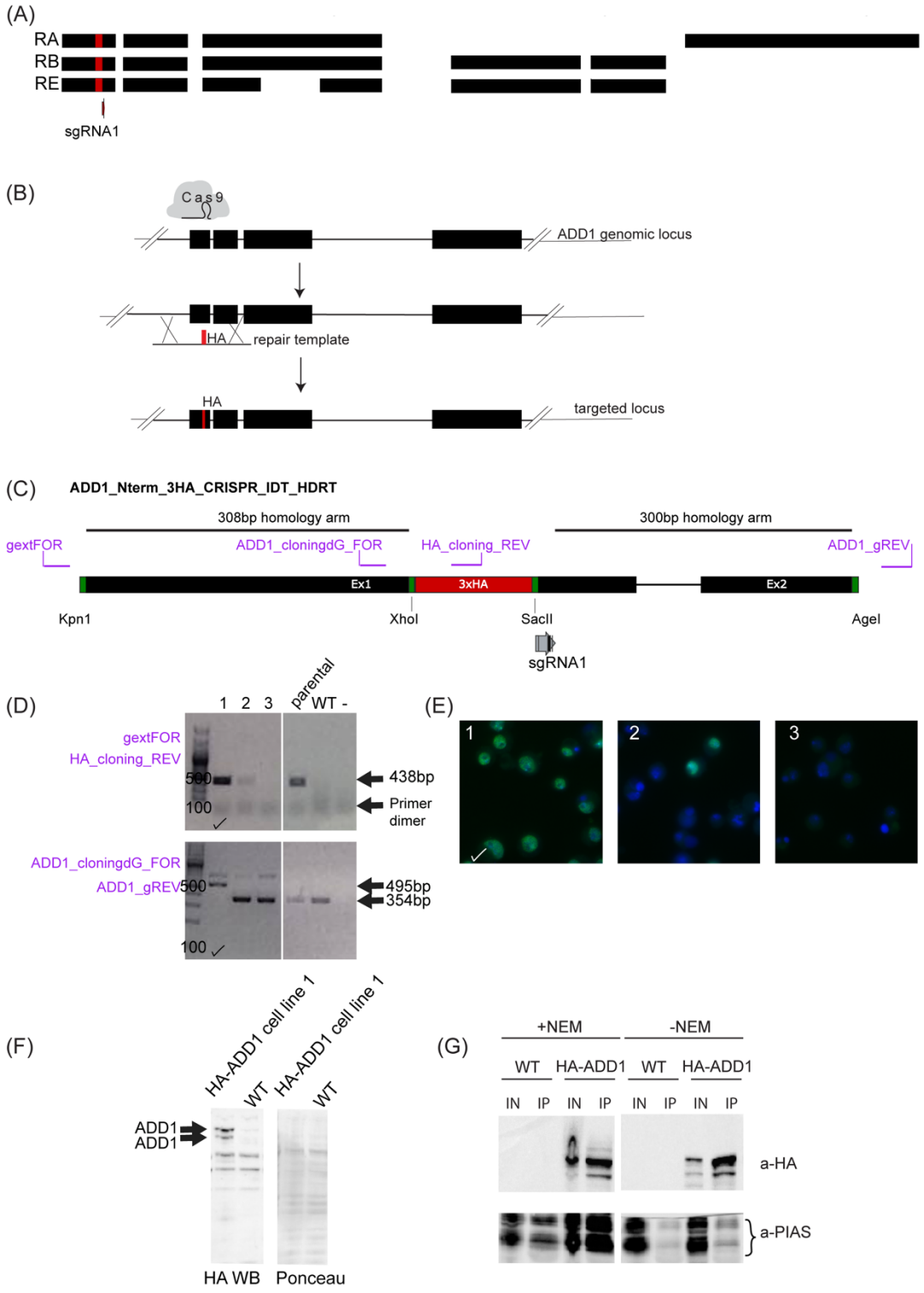

(A)
RA
RB
RE
sgRNA1
(B)
Cas9
ADD1 genomic locus
HA
repair template
HA
targeted locus
(C)
ADD1_Nterm_3HA_CRISPR_IDT_HDRT
308bp homology arm
300bp homology arm
gextFOR
ADD1_cloningdG_FOR
HA_cloning_REV
ADD1_gREV
Ex1
3xHA
Ex2
Kpn1
XhoI
SacII
AgeI
sgRNA1
(D)
1 2 3
parental
WT -
gextFOR
HA_cloning_REV
500
100
438bp
Primer dimer
ADD1_cloningdG_FOR
ADD1_gREV
500
100
495bp
354bp
(E)
1
2
3
(F)
HA-ADD1 cell line 1
WT
HA-ADD1 cell line 1
WT
ADD1
ADD1
HA WB
Ponceau
(G)
+NEM
-NEM
WT
HA-ADD1
WT
HA-ADD1
IN IP IN IP IN IP IN IP
a-HA
a-PIAS

Figure 8–11 Generation of HA-ADD1 S2 cell line

(A) Schematic of the three ADD1 isoforms, ADD1-RA, ADD1-RB- and ADD1-RE in *Drosophila* S2 cells. Exons are in black. The location of the sgRNA recognition site and tag are marked **(B)** Schematic of CRISPR/Cas9 insertion of an HA tag into the coding sequence of first exon of ADD1 **(C)** Design of the repair construct, 500bp IDT fragment designed for cloning to generate the repair template for integration of the HA tag. The length of the homology arms are shown. Marked in purple are the sites of primer binding. Restriction sites engineered for cloning and the site of sgRNA1 recognition are marked (lines in arrow show mutations to recognition sequence). **(D)** Clones were screened to confirm tag insertion at the ADD1 genomic locus by PCR using primers binding within the tagged region (HA_cloning_REV) and outside of the repair construct (gextFOR) resulting in a band of 438bp. If the repair construct didn't integrate, no DNA was amplified, and no band was detected. A second PCR using ADD1_cloningdG_FOR and ADD1_gREV was performed to confirm all four alleles of ADD1 were tagged by amplifying across the locus containing the insertion, resulting in two bands: without the tag (354bp) and with insertion (495bp). PCR products were run on 3% agarose and visualised by SybrSafe stain. **(E)** Immunofluorescence on settled S2 cells using an anti-HA mouse antibody recognising integration of the HA tag detected with anti-mouse AF 488. **(F)** HA-ADD1 cell line 1 S2 cells and wild-type (WT) S2 cells analysed by SDS-PAGE followed by western blot probing with anti-HA antibody, corresponding to tagged ADD1. Ponceau stain shows equal protein loading. **(G)** Immunoprecipitation of ADD1 from nuclear extracts generated from HA-ADD1 cell line 1 S2 cells and wild-type (WT) S2 cells analysed by SDS-PAGE followed by western blot probing with anti-HA antibody. IP washes were performed in 0.3M NaCl containing IP wash buffer. The membrane was cut at the 100kDa marker and the 100kDa to 50kDa region was probed with *Drosophila* anti-PIAS sheep antibody.

Appendix 5: MaxQuant SUMO Post-Translational Modification Identification

User	Cox
Name	QQTGG (K)
Description	SUMO-2/3 Q87R
Composition	H(29) C(18) N(7) O(8)
Position	Not C-term
Type	Standard
New terminus	N-term
Date last modified	14/10/2010
Specificities	[K,DP,b2-QQ,QQ,H(16) C(10) N(4) O(4),DP,b3-QQT,QQT,H(23) C(14) N(5) O(6),DP,b4-QQTG,QQTG,H(26) C(16) N(6) O(7),DP,b5-QQTGG,QQTGG,H(29) C(18) N(7) O(8),DP,b2-QQ-H2O,QQ-H2O,H(14) C(10) N(4) O(3),DP,b3-QQT-H2O,QQT-H2O,H(21) C(14) N(5) O(5),DP,b4-QQTG-H2O,QQTG-H2O,H(24) C(16) N(6) O(6),DP,H(27) C(18) N(7) O(7),H27 C18 N7 O7,H(27) C(18) N(7) O(7)]

Name	Short name	Composition	Neutral mass	M + H+
b2-QQ	QQ	H(16) C(10) N(4) O(4)	256.117155	257.1244315
b3-QQT	QQT	H(23) C(14) N(5) O(6)	357.1648335	358.17211
b4-QQTG	QQTG	H(26) C(16) N(6) O(7)	414.1862972	415.1935737
b5-QQTGG	**QQTGG**	**H(29) C(18) N(7) O(8)**	**471.2077609**	**472.2150374**
b2-QQ-H2O	QQ-H2O	H(14) C(10) N(4) O(3)	238.1065903	239.1138668
b3-QQT-H2O	QQT-H2O	H(21) C(14) N(5) O(5)	339.1542688	340.1615453
b4-QQTG-H2O	QQTG-H2O	H(24) C(16) N(6) O(6)	396.1757325	397.183009
H(27) C(18) N(7) O(7)	H27 C18 N7 O7	H(27) C(18) N(7) O(7)	453.1971963	454.2044727

Appendix 6: Published models influential to building the speculative models presented in Chapter 8.

Model 1 Five types of chromatin in *Drosophila*[57]

Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells

Guillaume J. Filion1,$, Joke G. van Bemmel1,$, Ulrich Braunschweig1,$, Wendy Talhout1, Jop Kind1, Lucas D. Ward2,3,4, Wim Brugman5, Ines de Castro Genebra de Jesus1,6, Ron M. Kerkhoven5, Harmen J. Bussemaker2,3, and Bas van Steensel1,*

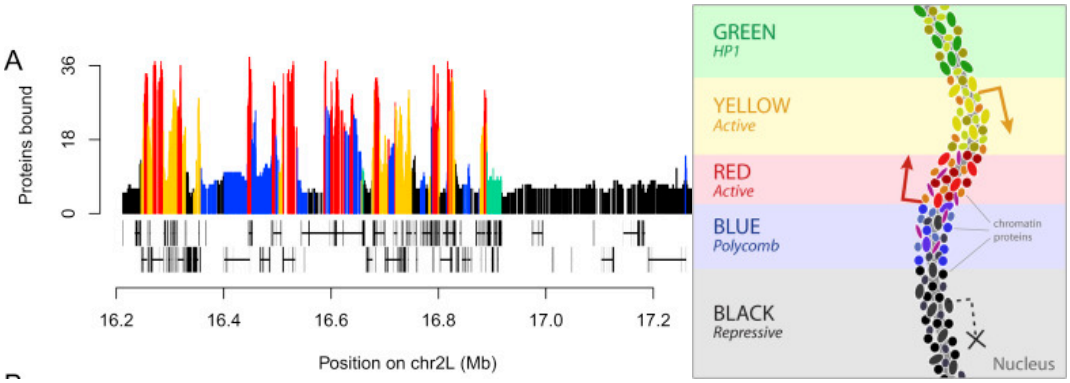


Figure S3: Chromatin Types Differ Widely in Their Total Protein Occupancy
(A) Sample plot of the total occupancy (out of 53 mapped proteins) on a 1 Mb segment of chromosome 2L. The height of each vertical line indicates the number of proteins bound to a locus. The color of the line indicates the local chromatin type.

Model 2 *Drosophila* chromatin compartment organisation[14]

Chromatin Architecture in the Fly: Living without CTCF/Cohesin Loop Extrusion? Alternating Chromatin States Provide a Basis for Domain Architecture in *Drosophila*

Nicholas E. Matthews and Rob White

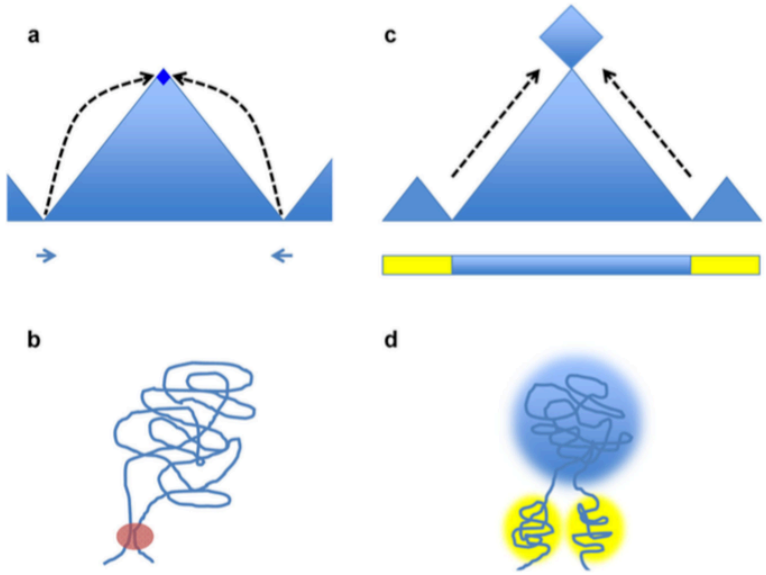


Figure 1. Domain topologies in mammals and *Drosophila*. a,b) In mammals many domains have a loop topology with the specific interaction of convergent CTCF-binding sites (blue arrows). a) Blue triangles indicate domains of enhanced interaction with an interaction "hotspot" at the apex and black dotted arrows indicate loop anchor interactions. b) Loop domain anchored by the CTCF/cohesin complex (red oval). c,d) In *Drosophila*, "inactive" domains (blue bar) are flanked by "active" domains (yellow bars). Black arrows in (c) indicate interaction between active domains, generating a "crown" of enhanced interaction above the inactive domain. d) Interaction between active domains forms a loose loop organization.

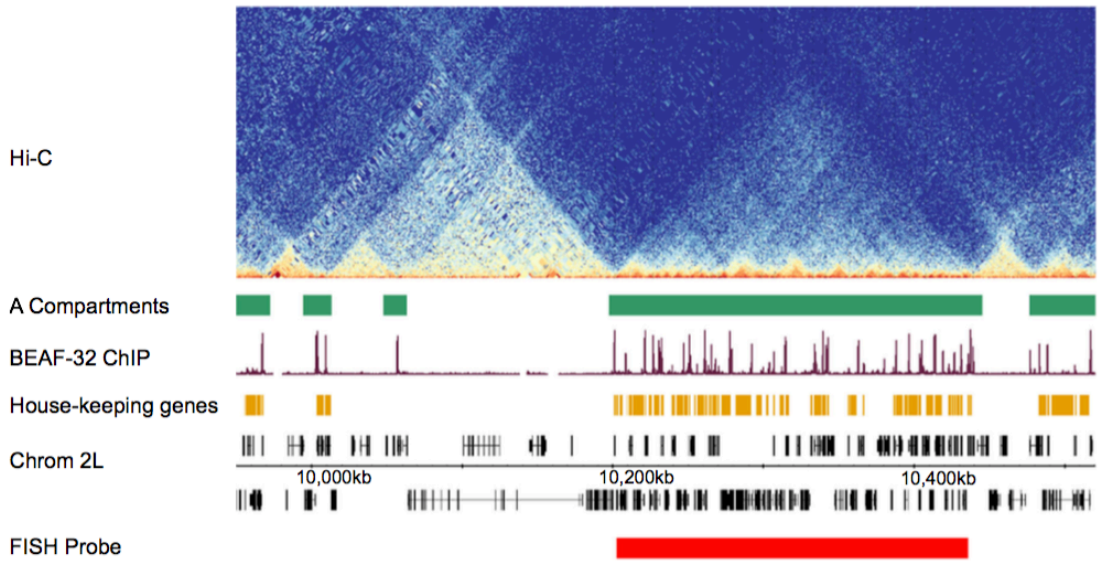
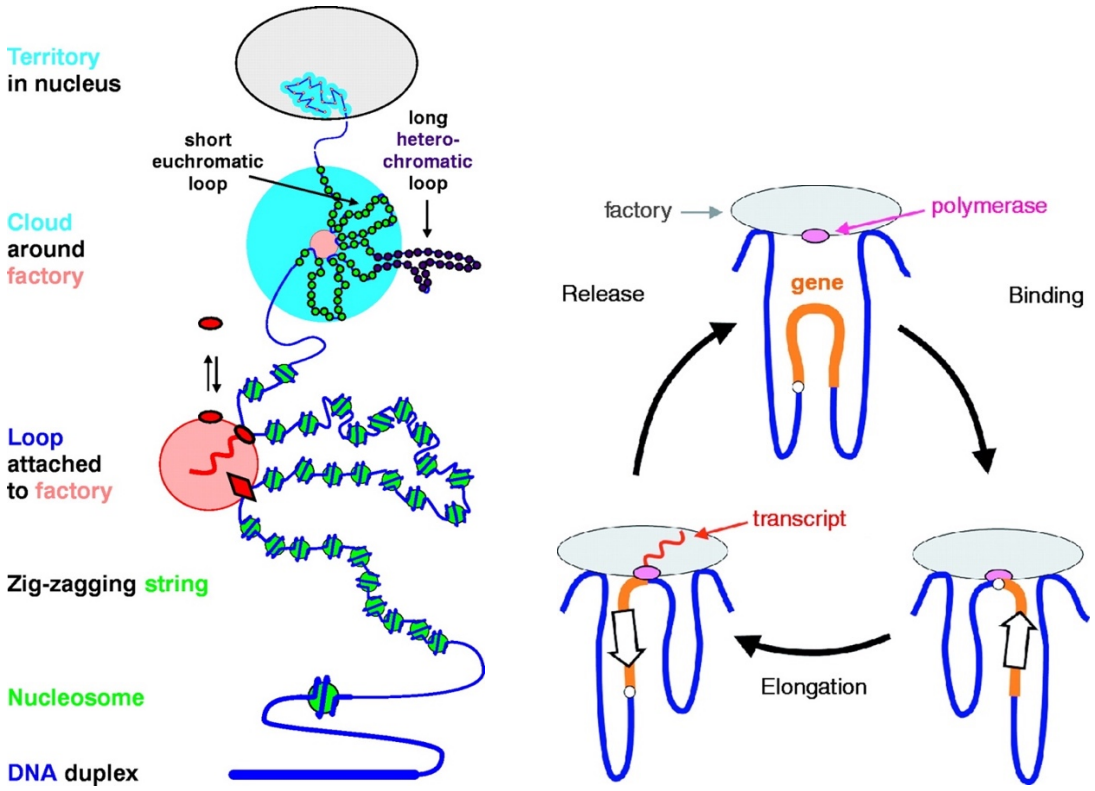


Figure 4. A compartments are enriched in insulator-binding proteins and housekeeping genes. Expanded view of the subregion from Figure 3a (chrom2L: 10 000 000–10 500 000). Red bar indicates the extent of the A compartment probe set visualized by in situ hybridization in Figure 6. This figure also illustrates different TAD hierarchies; the large TAD on the right, comprising an A compartment region, is further subdivided into minidomains (see Section 3.4 and Figure 5), whilst the large TAD on the left shows at least three levels of progressively larger domains, encompassing both A and B compartment regions. Hi-C from Kc167 cells,[32] A compartments from Kc167 cells,[21] BEAF-32 ChIP from Kc167 cells (ModENCODE GSM1535963), housekeeping genes from El-Sharnouby et al.[19] and the FISH probe from Szabo et al.[38] ChIP, chromatin immunoprecipitation.

Model 3 RNA Polymerase II transcription factories are clustered[347-349]

Hypotheses of Peter Cook



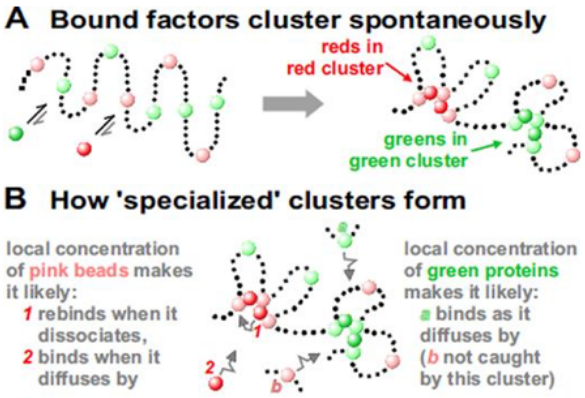


Figure 4: Model for the organization of the chromatin fibre in a HeLa nucleus. DNA is wound into a nucleosome, and then a zig-zagging string of nucleosomes is tied to a factory through a cluster of transcription factors (diamond) or an active polymerase (oval). Components of the factory exchange with the soluble pool, and attachments to the factory are made and broken as factors dissociate and transcription terminates. 10-20 loops (only three are shown) of 5-200 kbp form a cloud around the transcription factory; long, static, loops are likely to become heterochromatic and attached to the lamina. 50-100 clouds then form a chromosome territory. Modified, with permission, from Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc. (Cook, 2001).

Figure 5: A transcription cycle. A chromatin fiber is tied in loops (only one is shown) to a factory. The promoter (small circle) binds to one of the polymerases in the factory, and the transcript is generated as the template slides (open arrows) through the polymerase; at termination, the template detaches so the cycle can repeat. This model can be extended to explain how an inactive gene at the tip of a long heterochromatic loop could be activated (not shown). First, a transcription unit near the factory attaches, creating subloops. Then, the resulting transcription reels in the loop, “remodeling” and “opening” its chromatin. Now, other enhancers and transcription units attach, creating successively smaller loops until the inactive gene is brought sufficiently close to the factory to bind. Such transcription cycles can be incorporated into dynamic models for the way genomes are organized in prokaryotes and eukaryotes and for how chromosomes might pair during meiosis (51,52).

Figure 2: A process driving the spontaneous clustering of multivalent factors (a.k.a., the ‘bridging-induced attraction’). (A) Overview of one Brownian-dynamics simulation. Red and green ‘factors’ (colored spheres) bind reversibly to ‘chromatin’ (a string of beads); red factors bind only to pink beads, green factors only to light-green ones (non-binding beads shown as black dots). Bound factors spontaneously cluster—red with red and green with green—despite any specified interactions between proteins or between beads. (B) Explanation. Local concentrations create positive-feedback loops driving growth of nascent clusters; bound factors and binding beads rarely escape, and additional factors/beads are caught as they diffuse by. Red and green clusters are inevitably separate in 3D space because their cognate binding sites are separate in 1D sequence space. Cluster growth is limited by entropic costs of crowding together ever-more loops.

Model 4 Models of the fine-scale chromatin folding[360]

Resolving the 3D landscape of transcription-linked mammalian chromatin folding

Tsung-Han S. Hsieh, Elena Slobodyanyuk, Anders S. Hansen,
Claudia Cattoglio, Oliver J. Rando, Robert Tjian, Xavier Darzacq

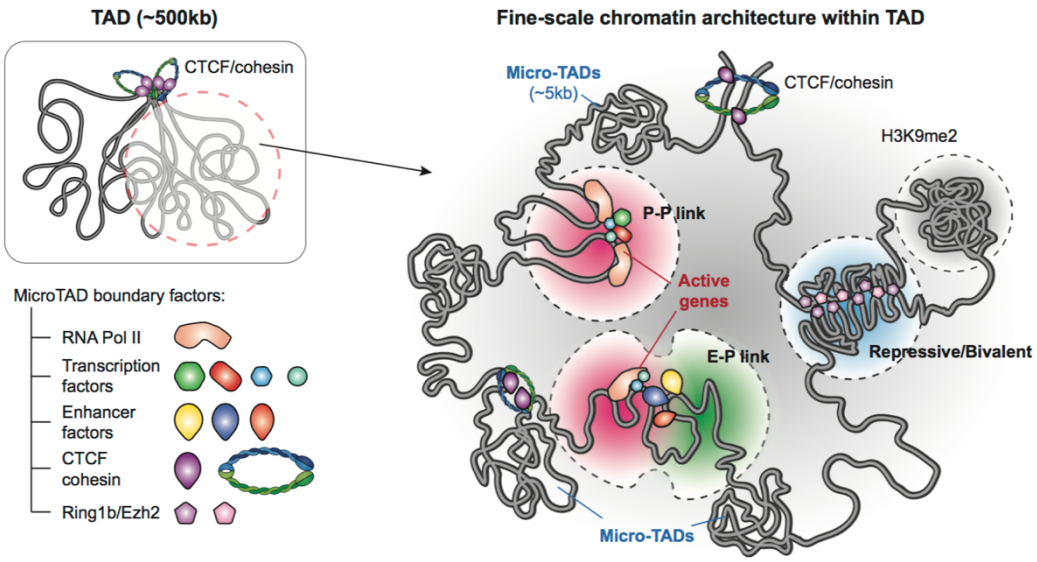


Fig. 6. (D) Schematics of the new layer of chromatin organization revealed by Micro-C. The fine-scale chromatin structures such microTADs, E-P/P-P links, and repressive bundle contacts build the basic units within TADs.

Enhancer-Promter (E-P)
Promoter-Promoter (P-P)

Model 5 SUMOylation of insulator elements (in *Drosophila*) in large-scale chromatin structure in interphase[226]

Victor Corces model

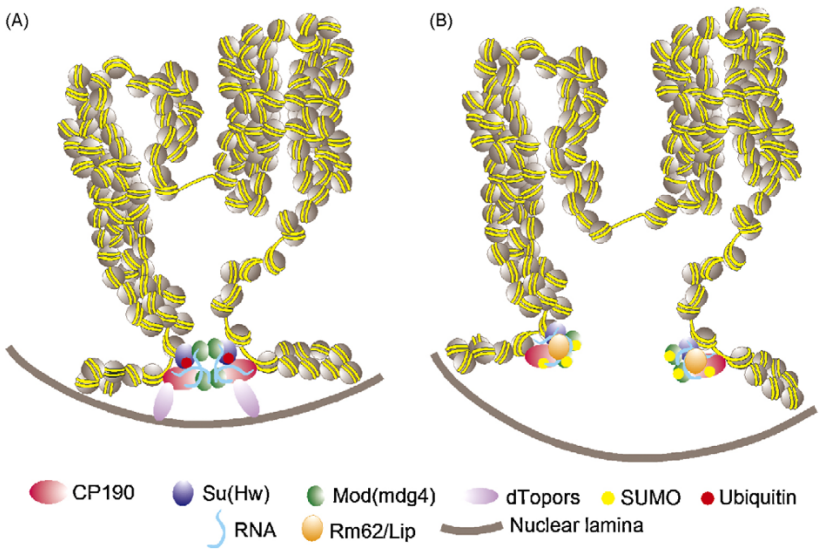
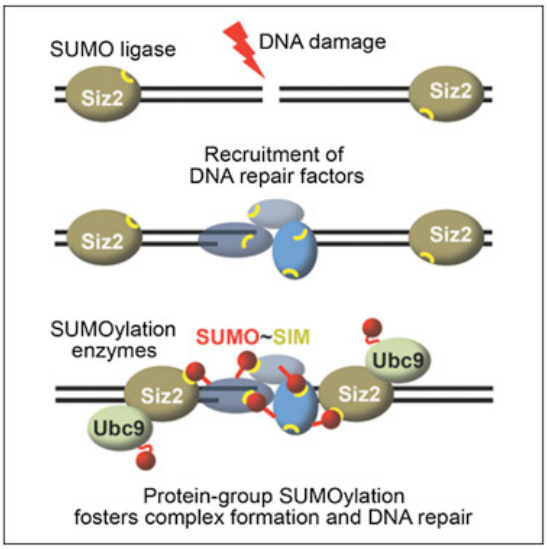


Fig. 2. Insulator activity can be regulated by ubiquitination and sumoylation of insulator proteins. (A) Two active insulators coming together at an insulator body. dTopors is present at the insulator sites, Rm62/Lip is not present, Su(Hw) is ubiquitinated, Mod(mdg4)2.2 and CP190 are not sumoylated and dTopors serves as a bridge to the lamina. (B) Two inactive insulators that cannot be part of an insulator body. dTopors is absent and Su(Hw) is not ubiquitinated, whereas Mod(mdg4)2.2 and CP190 are sumoylated. Rm62/Lip is present and bound to RNA. Under these conditions, the two insulator sites cannot interact and form insulator bodies. Absence of dTopors also precludes interactions with the lamina.

Model 6 "SUMO spray" [225]

Stefan Jentsch model



Graphical abstract: Protein modification by SUMO affects a wide range of protein substrates. Surprisingly, although SUMO pathway mutants display strong phenotypes, the function of individual SUMO modifications is often enigmatic, and SUMOylation-defective mutants commonly lack notable phenotypes. Here, we use DNA double-strand break repair as an example and show that DNA damage triggers a SUMOylation wave, leading to simultaneous multisite modifications of several repair proteins of the same pathway. Catalyzed by a DNA-bound SUMO ligase and triggered by single-stranded DNA, SUMOylation stabilizes physical interactions between the proteins. Notably, only wholesale elimination of SUMOylation of several repair proteins significantly affects the homologous recombination pathway by considerably slowing down DNA repair. Thus, SUMO acts synergistically on several proteins, and individual modifications only add up to efficient repair. We propose that SUMOylation may thus often target a protein group rather than individual proteins, whereas localized modification enzymes and highly specific triggers ensure specificity.

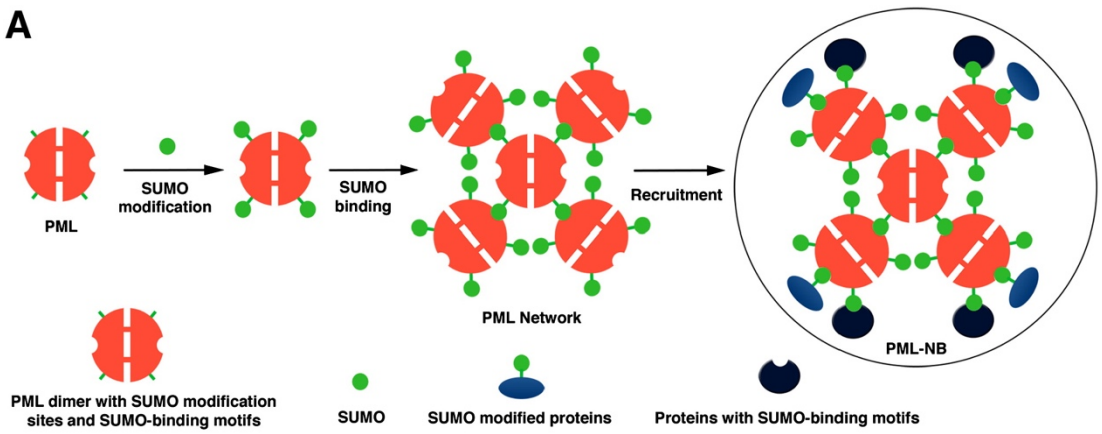
Model 7 SUMOylation of a protein that contains a SAP domain regulating transcription and pre-mRNA splicing[361]:

Jeffrey Parvin paper

Early steps of gene expression are a composite of promoter recognition, promoter activation, RNA synthesis and RNA processing, and it is known that SUMOylation, a post-translational modification, is involved in transcription regulation. We previously found that SUMO-1 marks chromatin at the proximal promoter regions of some of the most active housekeeping genes during interphase in human cells, but the SUMOylated targets on the chromatin remained unclear. In this study, we found that **SUMO-1 marks the promoters of ribosomal protein genes via modification of the Scaffold Associated Factor B (SAFB) protein**, and the **SUMOylated SAFB stimulated both the binding of RNA polymerase to promoters and pre-mRNA splicing.** Depletion of SAFB decreased RNA polymerase II binding to promoters and nuclear processing of the mRNA, though mRNA stability was not affected. This study reveals an unexpected role of SUMO-1 and SAFB in the stimulatory coupling of promoter binding, transcription initiation and RNA processing.

Model 8: SUMO: The Glue that Binds[350]

Michael J. Matunis, Xiang-Dong Zhang, Nathan A. Ellis



Covalent SUMO Modification and Noncovalent Interactions with SUMO through SUMO-Binding Motifs Mediate the Assembly of PML-NBs and the Dynamic Distribution of Proteins between PML-NBs and the Nucleoplasm

(A) PML forms a homodimer, with each subunit containing multiple SUMO modification sites (green lines) and a single SUMO-binding motif (notch). Intermolecular interactions, mediated by SUMO and SUMO-binding motifs, nucleate the formation of a PML network that is able to recruit an assortment of other SUMO-modified proteins and proteins with SUMO-binding motifs.

Appendix 6: Materials

Stock solutions

All water except where stated was Milli-Q (MQ). All stock solutions were autoclaved or filter sterilized where appropriate.

Ammonium bicarbonate (ABC): 50mM in water

Acetonitrile (ACN) 50mM in water

Ammonium persulphate (APS) prepared as 10% w/v solution in water, stored at 4ºC for 2 weeks or at -20ºC longer term

Antibiotics- 1000x solutions were prepared by dissolving the antibiotics listed in Table 17 in water. Aliquots were stored at -20ºC, with a working stock kept at 4ºC for a maximum of 2 weeks.

Table 17 Antibiotic stocks

Antibiotic	Powdered stock	Stock solution
Kanamycin	kanamycin monosulphate	50 mg/ml
Tetracycline	tetracycline hydrochloride	10 mg/ml
Ampicillin	ampicillin sodium salt	100 mg/ml
Carbenacillin	carbenicillin disodium	100 mg/ml
Gentamycin	gentamicin sulfate	10 µg/ml

Blasticidin (for S2 cell culture) 500mg/mL solution in water

Hygromycin B (for S2 cell culture) 100mg/mL solution in water

CHCA MALDI matrix saturated solution of 10mg a-cyano-4-hydroxycinnamic acid (CHCA), 400µL nuclease-free water, 100µL 1% (v/v) TFA and 500µL acetonitrile

Coomassie Coomassie brilliant blue 0.1% (w/v) in 50% (v/v) methanol, 10% (v/v) acetic acid in water.

Copper Sulfate ($CuSO_4$): 2mM solution in water

dNTPs 100mM dATP, dCTP, dGTP and dTTP (Fermentas) were mixed in equal ratios to obtain a 25mM stock of all four amino acids.

DAPI solution 4',6-diaminidino-2-phenylindole dissolved to a final concentration of 5µg/mL in 0.1% Triton X-100/PBS

Dithiothreitol (DTT) dissolved to final concentration 1.0M in water and stored in aliquots at -20°C

DNA loading buffer 5x TBE, 30% glycerol, 0.2% bromophenol blue in nuclease-free water (Ambicon)

DNA molecular weight marker Generuler 1 kb (Thermo Fisher Scientific) or 100 bp ladder (NEB) was dissolved at 500µg/mL in 1x DNA Loading Buffer.
500ng – 1µg was normally loaded per lane.

EDTA dissolved at 0.5M in water and adjusted to pH8.0 with NaOH

EGTA dissolved at 0.1M in water and adjusted to pH8.0 with NaOH

Glycerol 50% solution in water

HEPES solution 4-(2-hydroxyethyl)-1-piperazine-1- ethanesulfonic acid dissolved to final concentration 1.0M in water and pH adjusted to 8.0 with NaOH

Isopropyl β-D-1-thiogalactopyranoside (IPTG) dissolved to final concentration 1.0M in water and stored in aliquots at -20ºC

Iodoacetamide (IAA) 50mM stock, 50mM ABC in water

Lithium chloride (LiCl) 10M solution in water

Magnesium sulfate (MgSO4) 1M stock in water

Nuclease-free water (Ambicon)

Poly-Lysine 10% solution in water

Ponceau 0.1% w/w Ponceau S dye, 1% v/v acetic acid in water

Restriction enzymes and associated buffers were purchased from NEB

SDS 10% and 20% w/v stocks prepared in water

Sodium acetate (NaOAc) 0.6% (v/v) glacial acetic acid in 0.2M sodium acetate trihydrate in nuclease-free water, pH to 5.0 with 10N NaOH

Sodium chloride 5M stock in water

Trifluroacetic acid (TFA) 0.1% in water

Tris-buffers 2-amino-2(hydroxymethyl)-1,3-propanediol 1.0M solutions in water, pH adjusted with HCl or NaOH to make stocks at pH 6.0, 7.5 and 8.0

Triton X-100 10% solution in water

X-gal 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside dissolved to a final concentration of 20mg/mL in dimethyl sulfoxide (DMSO) and stored in small aliquots at -20ºC (light sensitive, so wrapped in foil to maintain light barrier)

General buffers

Cell lysis buffer 20mM Tris–HCl pH7.5, 20mM NaCl, 2mM $MgCl_2$, 1mM DTT, 1x Roche cOmplete™ EDTA- free Protease Inhibitors (Roche, 5056489001) supplemented with 20mM NEM where appropriate

Benzonase buffer 20mM Tris–HCl pH7.5, 20mM NaCl, 2mM $MgCl_2$, 1mM DTT, 1x Roche cOmplete™ EDTA- free Protease Inhibitors (Roche, 5056489001), 0.5% Triton X-100, 100U benzoase (Novagen)

Chromatin extraction buffer 20mM Tris–HCl pH7.5, 0.35M NaCl, 2mM $MgCl_2$, 1mM DTT, 1x Roche cOmplete™ EDTA- free Protease Inhibitors (Roche, 5056489001), 0.1% Triton X-100

Blocking buffer PBS, Tween20 0.1%, 3% Bovine Serum Albumine (BSA)

Enhancer Solution 11mg Coumaric Acid in 10mL Dimethyl sulfoxide (DMSO)

Embryo extraction buffer 20mM Tris–HCl pH7.5, 20mM NaCl, 2mM $MgCl_2$, 1mM DTT, 1x Roche cOmplete™ EDTA- free Protease Inhibitors (Roche, 5056489001), 0.1% Triton X-100

Embryo wash buffer 0.7% (w/v) NaCl, 0.05% (v/v) Triton X-100

Genomic lysis buffer 150 mM NaCl, 10mM EDTA pH 8.0, 0.5% SDS

GST purification buffer 20mM Tris–HCl pH7.5, 300mM NaCl, 2mM $MgCl_2$, 1mM DTT

GST purification buffer with glutathione 20mM Tris –HCl pH7.5, 300mM NaCl, 2mM $MgCl_2$, 1mM DTT, 20mM glutathione (added just before use) and adjusted to pH 8.0 with NaOH

High-salt buffer 20mM Tris–HCl pH7.5, 0.5M NaCl, 2mM $MgCl_2$, 1mM DTT, 1x Roche cOmplete™ EDTA- free Protease Inhibitors (Roche, 5056489001), 0.1% Triton X-100

HIS purification buffer (low imidazole) 20mM Tris–HCl pH8.0, 300mM NaCl, 2mM $MgCl_2$, 0.5mM B-mercaptoethanol, 10mM imidazole. (Note, B-mercaptoethanol was replaced with TCEP 1mM for final XNP purification).

HIS purification buffer (high imidazole) 20mM Tris pH8, 300mM NaCl, 2mM $MgCl_2$, 0.5mM β-Mercaptoethanol, 300mM imidazole (Note, B-mercaptoethanol was replaced with TCEP 1mM for final XNP purification).

HIS Wash Buffer 20mM Tris-HCl pH8, 300mM NaCl, 2mM $MgCl_2$, 0.5mM β-Mercaptoethanol, 30mM Imidazole (Note, B-mercaptoethanol was replaced with TCEP 1mM for final XNP purification).

Luminol Solution 60mg Luminol, 240mL 0.1M Tris-HCl pH 8.8, 74.4µL 30% H_2O_2

IF fixative 3.7% formaldehyde solution (Sigma) in 0.1% Triton X-100/PBS

IP buffer 20mM Tris–HCl pH7.5, 0.1M NaCl, 2mM $MgCl_2$, 1mM DTT, 1x Roche cOmplete™ EDTA- free Protease Inhibitors (Roche, 5056489001), 0.1% Triton X-100

IP wash buffer 20mM Tris–HCl pH7.5, 0.1, 0.3 or 0.5M NaCl, 2mM $MgCl_2$, 0.1% Triton X-100

Phosphate buffered saline (PBS) (prepared by the media kitchen) 137mM NaCl, 2.7mM KCl, 2mM KH2PO4, 10mM $Na_2HPO_4 \cdot 2H_2O$, pH 7.4

PBS-Tween20 0.1% 1X PBS with 0.1% v/v Tween-20

PBS-Triton 0.1% 1xPBS, 0.1% Triton X-100

PBTA 1% (w/v) bovine serum albumin, 0.02% (w/v) sodium azide, 0.05% (v/v) Triton X-100 in PBS

RIPA buffer 10mM Tris-HCl pH8, 1mM EDTA, 1% Triton, 0.1% sodium deoxycholate, 0.1% SDS, 140mM NaCl, 1x Roche cOmpleteTM EDTA- free Protease Inhibitors (Roche, 5056489001)

TFB1 30mM KAc,100mM RbCl, 10mM $CaCl_2*2H_2O$, 50mM $MnCl_2*4H_2O$, 15% glycerol, (adjusted to pH 5.8 with acetic acid)

TFB2 10mM MOPS, 10mM RbCl, 10mM $CaCl_2*2H_2O$, 15% glycerol, (adjusted to pH 6.5 with KOH)

Tris-acetate EDTA 1x (TAE) 40mM Tris, 20mM acetic acid, 1mM EDTA, pH 8.0

Tris-borate-EDTA 1x (TBE) 0.1M Tris base, 0.1M boric acid, 2mM EDTA (disodium salt)

Tris-EDTA 1x (TE) 10mM Tris-HCl, 1mM EDTA pH 8.0

Trypsin buffer 13ng/µL trypsin (Pierce™ Tryspin Protease, MS grade Thermo Fisher Scientific), 10mM ABC, 10% ACN

Tris-glycine running buffer 25mM Tris, 250mM glycine, 0.1% SDS

SDS-PAGE transfer buffer 25mM Tris, 192mM glycine, supplemented with 20% methanol before use and used cold

6x LB 60mM Tris pH 6.8, 0.6M DTT, 12% (w/v) SDS, 0.06% (w/v) bromophenol blue, 47% (v/v) glycerol

2x SDS sample buffer 125 mM Tris-HCl (pH 6.8), 4% (w/v) SDS, 20% (v/v) glycerol, 300 mM (10% v/v) β-mercaptoethanol, 0.1% bromophenol blue

Media

All solutions were prepared by the media kitchen.

Grape juice medium: 50g agar to 1L water and 0.6L grape juice.

Fly food: standard yeast-cornmeal agar medium (Ashburner, 1989[362])
(5g sugar, 60g maize meal, 20g dried yeast, 10g agar and 2.5g Nipagin to IL water).

Luria-Bertani (LB) broth: 10% NaCl, 10% bactotrytone and 5% yeast, sterilised by autoclaving and stored at RT

LB-agar: LB supplemented with 1.5% (w/v) agar. For plasmid selection on plates, antibiotics were added as required at concentrations listed in Table 17.

SOC 0.5% yeast extract, 2% tryptone, 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$, 20mM glucose

Bacterial strains

Stocks of bacterial strains used in this study were made and stored at -80ºC.

10beta was used for routine cloning and propagation of all plasmids

EmBAcy was used for transformation of plasmids for bacmid production

Antisera, antibodies, probes and beads

All primary antibodies used in this study are listed in Table 18 and secondary antibodies in Table 19.

Table 18 Primary antibodies used in this study.

WB- Western blot, IF- immunofluorescence, IP- immunoprecipitation. For IPs use at 3µg antibody/30µL Dynabeads™/200µg nuclear extract.

Antibody	species	Source, catalogue number, stock concentration	Application (dilution)
dPIAS	sheep	Ron Hay, Dundee, stock at 1.2mg/ml, polyclonal	WB (1:10000), IF (1:500) IP
Sheep anti pig IgG	sheep	Bethyl, 1 mg/ml	IP
dSmt3 (SUMO)	rabbit	Norbert Joswig, polyclonal	WB (1:5000), IF (1:1000)
Ubiquitin	mouse	P4D1, CST, 3936 monoclonal	WB (1:1000)
ATRX	rat	Andreas Thomae monoclonal, clone 2E10	WB (1:1000)
XNP	rabbit	ERF (this study), not quantified as not affinity purified	IF (1:500), WB (1:1000), IP (30µL serum)
ADD1	rabbit	ERF (this study) not quantified as not affinity purified	IF (1:1000) WB (1:5000), IP (9µL serum)
BEAF-32	mouse	Developmental Studies Hybridoma Bank at the University of Iowa (C1A9) 43ug/ml	WB (1:500) IP (1:40)
H3	rabbit	-	WB (1:1000)
Cenp-C	rabbit	Lenhner lab, polyclonal, Stephan Heidmann	WB (1:1000)
GST	rabbit	Sigma	WB 1:5000
Penta HIS	mouse	Qiagen 34660	WB 1:5000
FLAG	rabbit	Clone D6W58	WB (1:1000)
FLAG	mouse	M2 clone	WB (1:1000)
V5	mouse	Invitrogen, R96025	IF (1:100) WB (1:1000)

HA	mouse	Clone 12CA5, Simona Saccani (IRCAN, Nice)	IF (1:10000)
HP1a/ Su(var)205	Mouse	Barbara Wakimoto (C1A9) hybridoma supernatant	WB (1:1000)

Table 19 Secondary antibodies used in this study.

HRP – horseradish peroxidase, WB – Western blot, IF – immunofluorescence.

Antibody	species	Source, catalogue number, stock concentration	Application (dilution)
Donkey anti sheep HRP		-	WB (1:10000)
Mouse IgG HRP Linked Whole Ab		GE Healthcare, NA931-1ML	WB (1:10000)
Rabbit IgG HRP Linked Whole Ab		Sigma, GENA934-1ML	WB (1:10000)
Rat IgG, HRP-linked whole antibody		Sigma, GENA935	
Anti-rabbit 680RD	Rabbit	LICOR	WB (1:20000)
Anti-mouse FW800	mouse	LICOR	WB (1:20000)
Anti-rat FW800	rat	LICOR	WB (1:20000)
Anti-goat 680RD	goat	LICOR	WB (1:20000)
Alexa 488 F(ab) fragment goat anti-mouse	mouse	Invitrogen	IF (1:100)
Alexa 488 goat anti-rabbit	rabbit	Invitrogen	IF (1:100)
Alexa 555 donkey anti-sheep	sheep	Invitrogen	IF (1:100)
Alexa 647 F(ab) fragment goat anti-mouse	mouse	Invitrogen, A21237	IF (1:100)

His Probe HRP (Pierce 15165, 25 mg/ml, BSA TBS-T, 1:5000)

Fly line

Oregon-R-P2 or Oregon-R-C

Cell lines

Spodoptera frugiperda 9 (Sf9) cells in Sf-900™ II SFM, Catalog number: 11496015, Thermo Fisher Scientific (purchased April 2018)

Drosophila Schneider Line 2 derived L2–4 cell line (S2)
S2 ADD1^{HA} generated in this study
S2 ADD1^{HA} PIAS-C-V5 generated in this study
S2 PIAS-N-V5 generated in this study
S2 PIAS-C-V5 generated in this study

Plasmids used during this study
pMT_ATRX_FLAG (Heun lab)
pMT_Pias9_V5 (Heun lab)
pBlue (Heun lab)
pAc5.1-ADX-a-V5 (from Viviana Valadez-Graham[296])
pIB_Cas9_CD4_Blast
pFL (Cook lab)
pFastBac (Cook lab)

Plasmids generated during this study
pMT_PIASNter_V5_HDRT
pBlue_PIASC_V5_HDRT
pMT_ADD1_HA_HDRT
pFL_GST_ADD1
pFastBac XNP

Oligonucleotides

All primers and gene blocks were synthesized by Integrated DNA Technologies (IDT). Lyophilized Primers were resuspended in 1xTE to a concentration of 50μM and stored at -20°C as a stock solution. A 1:100 dilution was prepared as a working stock in double-distilled H_2O (ddH2O).

Table 20 List of primers

Purpose	Primer Name	Primer sequence	Tm	Date
To check insertion into the pFL or pFastBac vector	pFL&FastBac Fwd_seq	CCGGAATATTAATAGATCATGG	48.3	Cook lab stock
	pFL&FastBac_Rev_seq	CTCTAGATTCGAAAGCGGCC	55.6	
Sequencing of ADD1	ADD1RAseq3_F	GAGTAACAGTGCTCCGGGTAG	56.9	24-April-2017
	ADD1RAseq511_F	CCACTACGAGCCCAACACTG	58.0	24-April-2017
	ADD1RAseq1005_F	AGTGCCACCGACACGTAGTAG	58.7	24-April-2017
	ADD1RAseq1508_F	ATGCCAATGGCGGTCAAACTC	58.6	24-April-2017
	ADD1RAseq2000_F	AAGAACAGCCTCAGGAAACG	55.2	24-April-2017
	ADD1RAseq2502_F	ACACGGTGAGGGTGATAAG	54.1	24-April-2017
	ADD1RAseq3012_F	TGACAAGGAAGTGGAACCATC	54.8	24-April-2017
	ADD1RAseq3489_F	ACAAGTAGGAGCACCCAGC	57.2	24-April-2017
	ADD1RAseq3000_R	AGTGGAGCAGCAGTATACCTCGCC	62.4	24-April-2017
	ADD1RAseq6210_R	AACAATCTCAGCCGTGGCGAACTCG	63.2	24-April-2017

To confirm insertion of a tag at N-terminal PIAS core	PIAS_genomic_cloning_FOR PIAS_genomic_cloning_REV	TGTTCGCCATGCCATTTCCG CGTCCAGCGAATGAGATGTTCAG	58.9 57.9	27-FEB-2017 27-FEB-2017
To confirm insertion of a tag at C-terminal PIAS core	3'GSP2b PIASCter_gREV	CGACATGCCGCTGGCTAAG GGATGTATGTAAATGGGATGTGTATGG	59.0 55.5	16-Jun-2017 15-MAY-2017
To confirm insertion of a tag at N-terminal ADD1	gextFOR HA_cloning_REV ADD1_cloningdG_FOR ADD1_gREV	GCCTACTTCGGCTGGAACCCG GTCGGGGACGTCATACGGATAAGATCCGCC TCCAAGAGGCCCGGATACGGAACAG TGGGATGCATCTTGATGTTGC	62.6 65.8 64.0 56.0	06-APR-2017 30-Mar-2017 30-Mar-2017 30-Mar-2017
For amplification of sgRNAs from gene blocks containing a dU.6 promoter for use in CRISPR/Cas9 genome editing	dU6_2_sgRNA_F dU6_2_3_sgRNA_R	GTTCGACTTGCAGCCTGAAATACGGCACG AAAAAAGCACCGACTCGGTGCCACTTTTTCAA GTTGATAA	55.0	Heun lab stock

References

1 Vo Ngoc, L., Kassavetis, G. A. & Kadonaga, J. T. The RNA Polymerase II Core Promoter in *Drosophila*. *Genetics* **212**, 13-24, doi:10.1534/genetics.119.302021 (2019).

2 Kornberg, R. D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868-871 (1974).

3 Krude, T. Chromatin. Nucleosome assembly during DNA replication. *Curr Biol* **5**, 1232-1234 (1995).

4 Luger, K., Rechsteiner, T. J., Flaus, A. J., Waye, M. M. & Richmond, T. J. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J Mol Biol* **272**, 301-311, doi:10.1006/jmbi.1997.1235 (1997).

5 Berezney, R. & Coffey, D. S. Identification of a nuclear protein matrix. *Biochem Biophys Res Commun* **60**, 1410-1417, doi:10.1016/0006-291x(74)90355-6 (1974).

6 Jackson, D. A. & Cook, P. R. Replication occurs at a nucleoskeleton. *EMBO J* **5**, 1403-1410 (1986).

7 Jackson, D. A. & Cook, P. R. Transcription occurs at a nucleoskeleton. *EMBO J* **4**, 919-925 (1985).

8 Zeitlin, S., Parent, A., Silverstein, S. & Efstratiadis, A. Pre-mRNA splicing and the nuclear matrix. *Mol Cell Biol* **7**, 111-120, doi:10.1128/mcb.7.1.111 (1987).

9 Reyes, J. C., Muchardt, C. & Yaniv, M. Components of the human SWI/SNF complex are enriched in active chromatin and are associated with the nuclear matrix. *J Cell Biol* **137**, 263-274, doi:10.1083/jcb.137.2.263 (1997).

10 Nozawa, R. S. & Gilbert, N. RNA: Nuclear Glue for Folding the Genome. *Trends Cell Biol* **29**, 201-211, doi:10.1016/j.tcb.2018.12.003 (2019).

11 Li, X. & Fu, X. D. Chromatin-associated RNAs as facilitators of functional genomic interactions. *Nat Rev Genet*, doi:10.1038/s41576-019-0135-1 (2019).

12 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).

13 Rowley, M. J. *et al.* Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol Cell* **67**, 837-852.e837, doi:10.1016/j.molcel.2017.07.022 (2017).

14 Matthews, N. E. & White, R. Chromatin Architecture in the Fly: Living without CTCF/Cohesin Loop Extrusion?: Alternating Chromatin States Provide a Basis for Domain Architecture in Drosophila. *Bioessays*, e1900048, doi:10.1002/bies.201900048 (2019).

15 Li, L. *et al.* Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. *Mol Cell* **58**, 216-231, doi:10.1016/j.molcel.2015.02.023 (2015).

16 Hou, C., Li, L., Qin, Z. S. & Corces, V. G. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol Cell* **48**, 471-484, doi:10.1016/j.molcel.2012.08.031 (2012).

17 Stadler, M. R., Haines, J. E. & Eisen, M. B. Convergence of topological domain boundaries, insulators, and polytene interbands revealed by high-resolution mapping of chromatin contacts in the early. *Elife* **6**, doi:10.7554/eLife.29550 (2017).
18 Brackley, C. A., Taylor, S., Papantonis, A., Cook, P. R. & Marenduzzo, D. Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization. *Proc Natl Acad Sci U S A* **110**, E3605-3611, doi:10.1073/pnas.1302950110 (2013).
19 Mitchell, J. A. & Fraser, P. Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev* **22**, 20-25, doi:10.1101/gad.454008 (2008).
20 Robson, M. I., Ringel, A. R. & Mundlos, S. Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D. *Mol Cell* **74**, 1110-1122, doi:10.1016/j.molcel.2019.05.032 (2019).
21 Erdel, F. & Rippe, K. Formation of Chromatin Subcompartments by Phase Separation. *Biophys J* **114**, 2262-2270, doi:10.1016/j.bpj.2018.03.011 (2018).
22 Hansen, J. C. *et al.* The 10-nm chromatin fiber and its relationship to interphase chromosome organization. *Biochem Soc Trans* **46**, 67-76, doi:10.1042/BST20170101 (2018).
23 Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41-45, doi:10.1038/47412 (2000).
24 Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705, doi:10.1016/j.cell.2007.02.005 (2007).
25 Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res* **21**, 381-395, doi:10.1038/cr.2011.22 (2011).
26 Campos, E. I. & Reinberg, D. Histones: annotating chromatin. *Annu Rev Genet* **43**, 559-599, doi:10.1146/annurev.genet.032608.103928 (2009).
27 Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257 (1999).
28 Morgan, H. D., Santos, F., Green, K., Dean, W. & Reik, W. Epigenetic reprogramming in mammals. *Hum Mol Genet* **14 Spec No 1**, R47-58, doi:10.1093/hmg/ddi114 (2005).
29 Jaenisch, R. DNA methylation and imprinting: why bother? *Trends Genet* **13**, 323-329 (1997).
30 Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* **9**, 465-476, doi:10.1038/nrg2341 (2008).
31 Bird, A. P. CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209-213, doi:10.1038/321209a0 (1986).
32 Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L. & Bird, A. P. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* **58**, 499-507 (1989).
33 Jones, P. L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* **19**, 187-191, doi:10.1038/561 (1998).
34 Nan, X. *et al.* Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. *Proc Natl Acad Sci U S A* **104**, 2709-2714, doi:10.1073/pnas.0608056104 (2007).

35 Zhang, Y. *et al.* Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev* **13**, 1924-1935, doi:10.1101/gad.13.15.1924 (1999).

36 Fuks, F., Burgers, W. A., Godin, N., Kasai, M. & Kouzarides, T. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *EMBO J* **20**, 2536-2544, doi:10.1093/emboj/20.10.2536 (2001).

37 Rajavelu, A. *et al.* Chromatin-dependent allosteric regulation of DNMT3A activity by MeCP2. *Nucleic Acids Res* **46**, 9044-9056, doi:10.1093/nar/gky715 (2018).

38 Cardoso, C. *et al.* Specific interaction between the XNP/ATR-X gene product and the SET domain of the human EZH2 protein. *Hum Mol Genet* **7**, 679-684 (1998).

39 Sarma, K. *et al.* ATRX directs binding of PRC2 to Xist RNA and Polycomb targets. *Cell* **159**, 869-883, doi:10.1016/j.cell.2014.10.019 (2014).

40 Deshmukh, S., Ponnaluri, V. C., Dai, N., Pradhan, S. & Deobagkar, D. Levels of DNA cytosine methylation in the *Drosophila* genome. *PeerJ* **6**, e5119, doi:10.7717/peerj.5119 (2018).

41 Capuano, F., Mülleder, M., Kok, R., Blom, H. J. & Ralser, M. Cytosine DNA methylation is found in Drosophila melanogaster but absent in Saccharomyces cerevisiae, Schizosaccharomyces pombe, and other yeast species. *Anal Chem* **86**, 3697-3702, doi:10.1021/ac500447w (2014).

42 Bewick, A. J., Vogel, K. J., Moore, A. J. & Schmitz, R. J. Evolution of DNA Methylation across Insects. *Mol Biol Evol* **34**, 654-665, doi:10.1093/molbev/msw264 (2017).

43 Kunert, N., Marhold, J., Stanke, J., Stach, D. & Lyko, F. A Dnmt2-like protein mediates DNA methylation in Drosophila. *Development* **130**, 5083-5090, doi:10.1242/dev.00716 (2003).

44 Rudolph, T. *et al.* Heterochromatin formation in Drosophila is initiated through active removal of H3K4 methylation by the LSD1 homolog SU(VAR)3-3. *Mol Cell* **26**, 103-115, doi:10.1016/j.molcel.2007.02.025 (2007).

45 Weber, C. M. & Henikoff, S. Histone variants: dynamic punctuation in transcription. *Genes Dev* **28**, 672-682, doi:10.1101/gad.238873.114 (2014).

46 Wong, L. H. *et al.* ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. *Genome Res* **20**, 351-360, doi:10.1101/gr.101477.109 (2010).

47 Goldberg, A. D. *et al.* Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* **140**, 678-691, doi:10.1016/j.cell.2010.01.003 (2010).

48 Bachu, M. *et al.* A versatile mouse model of epitope-tagged histone H3.3 to study epigenome dynamics. *J Biol Chem* **294**, 1904-1914, doi:10.1074/jbc.RA118.005550 (2019).

49 Tang, M. C. *et al.* Contribution of the two genes encoding histone variant h3.3 to viability and fertility in mice. *PLoS Genet* **11**, e1004964, doi:10.1371/journal.pgen.1004964 (2015).

50 Szenker, E., Ray-Gallet, D. & Almouzni, G. The double face of the histone variant H3.3. *Cell Res* **21**, 421-434, doi:10.1038/cr.2011.14 (2011).

51 Voon, H. P. & Wong, L. H. New players in heterochromatin silencing: histone variant H3.3 and the ATRX/DAXX chaperone. *Nucleic Acids Res* **44**, 1496-1501, doi:10.1093/nar/gkw012 (2016).

52 Jang, C. W., Shibata, Y., Starmer, J., Yee, D. & Magnuson, T. Histone H3.3 maintains genome integrity during mammalian development. *Genes Dev* **29**, 1377-1392, doi:10.1101/gad.264150.115 (2015).
53 Lin, C. J., Conti, M. & Ramalho-Santos, M. Histone variant H3.3 maintains a decondensed chromatin state essential for mouse preimplantation development. *Development* **140**, 3624-3634, doi:10.1242/dev.095513 (2013).
54 Heitz, E. Vol. 69 762-818 (I Jb Wiss Bot, 1928).
55 Comings, D. E. Arrangement of chromatin in the nucleus. *Hum Genet* **53**, 131-143 (1980).
56 Boldyreva, L. V. *et al.* Protein and Genetic Composition of Four Chromatin Types in *Drosophila melanogaster* Cell Lines. *Curr Genomics* **18**, 214-226, doi:10.2174/1389202917666160512164913 (2017).
57 Filion, G. J. *et al.* Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell* **143**, 212-224, doi:10.1016/j.cell.2010.09.009 (2010).
58 Kharchenko, P. V. *et al.* Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**, 480-485, doi:10.1038/nature09725 (2011).
59 Roy, S. *et al.* Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science* **330**, 1787-1797, doi:10.1126/science.1198374 (2010).
60 Peacock, W. J. *et al.* Fine structure and evolution of DNA in heterochromatin. *Cold Spring Harb Symp Quant Biol* **42 Pt 2**, 1121-1135 (1978).
61 Schotta, G. *et al.* Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *EMBO J* **21**, 1121-1131, doi:10.1093/emboj/21.5.1121 (2002).
62 Eissenberg, J. C., Morris, G. D., Reuter, G. & Hartnett, T. The heterochromatin-associated protein HP-1 is an essential protein in Drosophila with dosage-dependent effects on position-effect variegation. *Genetics* **131**, 345-352 (1992).
63 Janssen, A., Colmenares, S. U. & Karpen, G. H. Heterochromatin: Guardian of the Genome. *Annu Rev Cell Dev Biol* **34**, 265-288, doi:10.1146/annurev-cellbio-100617-062653 (2018).
64 Udugama, M. *et al.* Histone variant H3.3 provides the heterochromatic H3 lysine 9 tri-methylation mark at telomeres. *Nucleic Acids Res* **43**, 10227-10237, doi:10.1093/nar/gkv847 (2015).
65 Penke, T. J. R., McKay, D. J., Strahl, B. D., Matera, A. G. & Duronio, R. J. Functional Redundancy of Variant and Canonical Histone H3 Lysine 9 Modification in. *Genetics* **208**, 229-244, doi:10.1534/genetics.117.300480 (2018).
66 Avivi, S. *et al.* Visualizing nuclear RNAi activity in single living human cells. *Proc Natl Acad Sci U S A* **114**, E8837-E8846, doi:10.1073/pnas.1707440114 (2017).
67 Johnson, W. L. *et al.* RNA-dependent stabilization of SUV39H1 at constitutive heterochromatin. *Elife* **6**, doi:10.7554/eLife.25299 (2017).
68 Allshire, R. C. & Madhani, H. D. Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol* **19**, 229-244, doi:10.1038/nrm.2017.119 (2018).

69 Saksouk, N., Simboeck, E. & Déjardin, J. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin* **8**, 3, doi:10.1186/1756-8935-8-3 (2015).

70 Pal-Bhadra, M. *et al.* Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. *Science* **303**, 669-672, doi:10.1126/science.1092653 (2004).

71 Jih, G. *et al.* Unique roles for histone H3K9me states in RNAi and heritable silencing of transcription. *Nature* **547**, 463-467, doi:10.1038/nature23267 (2017).

72 Lin, H. & Yin, H. A novel epigenetic mechanism in Drosophila somatic cells mediated by Piwi and piRNAs. *Cold Spring Harb Symp Quant Biol* **73**, 273-281, doi:10.1101/sqb.2008.73.056 (2008).

73 Wang, S. H. & Elgin, S. C. Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. *Proc Natl Acad Sci U S A* **108**, 21164-21169, doi:10.1073/pnas.1107892109 (2011).

74 Cox, D. N. *et al.* A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev* **12**, 3715-3727, doi:10.1101/gad.12.23.3715 (1998).

75 Kalmykova, A. I., Klenov, M. S. & Gvozdev, V. A. Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline. *Nucleic Acids Res* **33**, 2052-2059, doi:10.1093/nar/gki323 (2005).

76 Teixeira, F. K. *et al.* piRNA-mediated regulation of transposon alternative splicing in the soma and germ line. *Nature* **552**, 268-272, doi:10.1038/nature25018 (2017).

77 Saito, K. *et al.* Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome. *Genes Dev* **20**, 2214-2222, doi:10.1101/gad.1454806 (2006).

78 Gu, T. & Elgin, S. C. Maternal depletion of Piwi, a component of the RNAi system, impacts heterochromatin formation in Drosophila. *PLoS Genet* **9**, e1003780, doi:10.1371/journal.pgen.1003780 (2013).

79 Rojas-Ríos, P., Chartier, A., Pierson, S. & Simonelig, M. Aubergine and piRNAs promote germline stem cell self-renewal by repressing the proto-oncogene. *EMBO J* **36**, 3194-3211, doi:10.15252/embj.201797259 (2017).

80 Ozata, D. M., Gainetdinov, I., Zoch, A., O'Carroll, D. & Zamore, P. D. PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet* **20**, 89-108, doi:10.1038/s41576-018-0073-3 (2019).

81 Aravin, A. A. *et al.* A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* **31**, 785-799, doi:10.1016/j.molcel.2008.09.003 (2008).

82 Kuramochi-Miyagawa, S. *et al.* DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* **22**, 908-917, doi:10.1101/gad.1640708 (2008).

83 Watanabe, T. *et al.* Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus. *Science* **332**, 848-852, doi:10.1126/science.1203919 (2011).

84 Castel, S. E. & Martienssen, R. A. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat Rev Genet* **14**, 100-112, doi:10.1038/nrg3355 (2013).

85 Yuan, K. & O'Farrell, P. H. TALE-light imaging reveals maternally guided, H3K9me2/3-independent emergence of functional heterochromatin in Drosophila embryos. *Genes Dev* **30**, 579-593, doi:10.1101/gad.272237.115 (2016).
86 Brower-Toland, B. *et al.* Drosophila PIWI associates with chromatin and interacts directly with HP1a. *Genes Dev* **21**, 2300-2311, doi:10.1101/gad.1564307 (2007).
87 Mendez, D. L., Mandt, R. E. & Elgin, S. C. Heterochromatin Protein 1a (HP1a) partner specificity is determined by critical amino acids in the chromo shadow domain and C-terminal extension. *J Biol Chem* **288**, 22315-22323, doi:10.1074/jbc.M113.468413 (2013).
88 Lo, P. K. *et al.* RNA helicase Belle/DDX3 regulates transgene expression in Drosophila. *Dev Biol* **412**, 57-70, doi:10.1016/j.ydbio.2016.02.014 (2016).
89 Shpiz, S., Ryazansky, S., Olovnikov, I., Abramov, Y. & Kalmykova, A. Euchromatic transposon insertions trigger production of novel Pi- and endo-siRNAs at the target sites in the drosophila germline. *PLoS Genet* **10**, e1004138, doi:10.1371/journal.pgen.1004138 (2014).
90 Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R. & Hannon, G. J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146-1150, doi:10.1126/science.1064023 (2001).
91 Caudy, A. A., Myers, M., Hannon, G. J. & Hammond, S. M. Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes Dev* **16**, 2491-2496, doi:10.1101/gad.1025202 (2002).
92 Lee, S. K. *et al.* Topoisomerase 3β interacts with RNAi machinery to promote heterochromatin formation and transcriptional silencing in Drosophila. *Nat Commun* **9**, 4946, doi:10.1038/s41467-018-07101-4 (2018).
93 Bozzetti, M. P. *et al.* The Drosophila fragile X mental retardation protein participates in the piRNA pathway. *J Cell Sci* **128**, 2070-2084, doi:10.1242/jcs.161810 (2015).
94 Cho, S., Park, J. S. & Kang, Y. K. AGO2 and SETDB1 cooperate in promoter-targeted transcriptional silencing of the androgen receptor gene. *Nucleic Acids Res* **42**, 13545-13556, doi:10.1093/nar/gku788 (2014).
95 Deshpande, G., Calhoun, G. & Schedl, P. Drosophila argonaute-2 is required early in embryogenesis for the assembly of centric/centromeric heterochromatin, nuclear division, nuclear migration, and germ-cell formation. *Genes Dev* **19**, 1680-1685, doi:10.1101/gad.1316805 (2005).
96 Porro, A. *et al.* Functional characterization of the TERRA transcriptome at damaged telomeres. *Nat Commun* **5**, 5379, doi:10.1038/ncomms6379 (2014).
97 Shirai, A. *et al.* Impact of nucleic acid and methylated H3K9 binding activities of Suv39h1 on its heterochromatin assembly. *Elife* **6**, doi:10.7554/eLife.25317 (2017).
98 Velazquez Camacho, O. *et al.* Major satellite repeat RNA stabilize heterochromatin retention of Suv39h enzymes by RNA-nucleosome association and RNA:DNA hybrid formation. *Elife* **6**, doi:10.7554/eLife.25293 (2017).
99 Maison, C., Quivy, J. P. & Almouzni, G. Suv39h1 links the SUMO pathway to constitutive heterochromatin. *Mol Cell Oncol* **3**, e1225546, doi:10.1080/23723556.2016.1225546 (2016).

100 Tschiersch, B. *et al.* The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes. *EMBO J* **13**, 3822-3831 (1994).
101 Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116-120, doi:10.1038/35065132 (2001).
102 Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120-124, doi:10.1038/35065138 (2001).
103 Larson, A. G. *et al.* Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236-240, doi:10.1038/nature22822 (2017).
104 Schotta, G. *et al.* A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev* **18**, 1251-1262, doi:10.1101/gad.300704 (2004).
105 Fuks, F., Hurd, P. J., Deplus, R. & Kouzarides, T. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res* **31**, 2305-2312, doi:10.1093/nar/gkg332 (2003).
106 Zhang, C. L., McKinsey, T. A. & Olson, E. N. Association of class II histone deacetylases with heterochromatin protein 1: potential role for histone methylation in control of muscle differentiation. *Mol Cell Biol* **22**, 7302-7312, doi:10.1128/mcb.22.20.7302-7312.2002 (2002).
107 Danzer, J. R. & Wallrath, L. L. Mechanisms of HP1-mediated gene silencing in Drosophila. *Development* **131**, 3571-3580, doi:10.1242/dev.01223 (2004).
108 Schneiderman, J. I., Goldstein, S. & Ahmad, K. Perturbation analysis of heterochromatin-mediated gene silencing and somatic inheritance. *PLoS Genet* **6**, e1001095, doi:10.1371/journal.pgen.1001095 (2010).
109 Savitsky, M., Kwon, D., Georgiev, P., Kalmykova, A. & Gvozdev, V. Telomere elongation is under the control of the RNAi-based mechanism in the Drosophila germline. *Genes Dev* **20**, 345-354, doi:10.1101/gad.370206 (2006).
110 Groth, A., Rocha, W., Verreault, A. & Almouzni, G. Chromatin challenges during DNA replication and repair. *Cell* **128**, 721-733, doi:10.1016/j.cell.2007.01.030 (2007).
111 Saha, A., Wittmeyer, J. & Cairns, B. R. Chromatin remodelling: the industrial revolution of DNA around histones. *Nat Rev Mol Cell Biol* **7**, 437-447, doi:10.1038/nrm1945 (2006).
112 Martens, J. A. & Winston, F. Recent advances in understanding chromatin remodeling by Swi/Snf complexes. *Curr Opin Genet Dev* **13**, 136-142 (2003).
113 Flaus, A., Martin, D. M., Barton, G. J. & Owen-Hughes, T. Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Res* **34**, 2887-2905, doi:10.1093/nar/gkl295 (2006).
114 Talbert, P. B. & Henikoff, S. Histone variants--ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol* **11**, 264-275, doi:10.1038/nrm2861 (2010).
115 Hoelper, D., Huang, H., Jain, A. Y., Patel, D. J. & Lewis, P. W. Structural and mechanistic insights into ATRX-dependent and -independent functions of the histone chaperone DAXX. *Nat Commun* **8**, 1193, doi:10.1038/s41467-017-01206-y (2017).

116 Lewis, P. W., Elsaesser, S. J., Noh, K. M., Stadler, S. C. & Allis, C. D. Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proc Natl Acad Sci U S A* **107**, 14075-14080, doi:10.1073/pnas.1008850107 (2010).
117 Morozov, V. M., Giovinazzi, S. & Ishov, A. M. CENP-B protects centromere chromatin integrity by facilitating histone deposition via the H3.3-specific chaperone Daxx. *Epigenetics Chromatin* **10**, 63, doi:10.1186/s13072-017-0164-y (2017).
118 Elgin, S. C. & Reuter, G. Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. *Cold Spring Harb Perspect Biol* **5**, a017780, doi:10.1101/cshperspect.a017780 (2013).
119 Li, X. Y., Harrison, M. M., Villalta, J. E., Kaplan, T. & Eisen, M. B. Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition. *Elife* **3**, doi:10.7554/eLife.03737 (2014).
120 Foe, V. E., Odell, G. M. & Edgar, V. A. *Mitosis and morphogenesis in the Drosophila embryo: Point and counterpoint*. 149-300 (1994).
121 Jachowicz, J. W., Santenard, A., Bender, A., Muller, J. & Torres-Padilla, M. E. Heterochromatin establishment at pericentromeres depends on nuclear position. *Genes Dev* **27**, 2427-2432, doi:10.1101/gad.224550.113 (2013).
122 Yuan, K., Seller, C. A., Shermoen, A. W. & O'Farrell, P. H. Timing the Drosophila Mid-Blastula Transition: A Cell Cycle-Centered View. *Trends Genet* **32**, 496-507, doi:10.1016/j.tig.2016.05.006 (2016).
123 Seller, C. A., Cho, C. Y. & O'Farrell, P. H. Rapid embryonic cell cycles defer the establishment of heterochromatin by Eggless/SetDB1 in. *Genes Dev* **33**, 403-417, doi:10.1101/gad.321646.118 (2019).
124 Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547**, 241-245, doi:10.1038/nature22989 (2017).
125 Armstrong, R. L. & Duronio, R. J. Phasing in heterochromatin during development. *Genes Dev* **33**, 379-381, doi:10.1101/gad.324731.119 (2019).
126 Rogers, A. K. *Mechanisms of Transcriptional Silencing by the Nuclear Piwi Protein in Drosophila Germ Cells* Doctor of Philosophy in Biology thesis, CALIFORNIA INSTITUTE OF TECHNOLOGY, (2018).
127 Kotadia, S., Crest, J., Tram, U., Riggs, B. & Sullivan, W. (eLS. John Wiley & Sons, Ltd, 2001).
128 Muller, H. Types of visible variations induced by X-rays in *Drosophila*. *Journal of Genetics* **22**, 299-334 (1930).
129 Locke, J., Kotarski, M. A. & Tartof, K. D. Dosage-dependent modifiers of position effect variegation in Drosophila and a mass action model that explains their effect. *Genetics* **120**, 181-198 (1988).
130 Schotta, G., Ebert, A., Dorn, R. & Reuter, G. Position-effect variegation and the genetic dissection of chromatin regulation in Drosophila. *Semin Cell Dev Biol* **14**, 67-75 (2003).
131 Reuter, G. & Wolff, I. Isolation of dominant suppressor mutations for position-effect variegation in Drosophila melanogaster. *Mol Gen Genet* **182**, 516-519 (1981).
132 Fromental-Ramain, C., Ramain, P. & Hamiche, A. The Drosophila DAXX-Like Protein (DLP) Cooperates with ASF1 for H3.3 Deposition and Heterochromatin Formation. *Mol Cell Biol* **37**, doi:10.1128/MCB.00597-16 (2017).

133 Mottus, R., Sobel, R. E. & Grigliatti, T. A. Mutational analysis of a histone deacetylase in Drosophila melanogaster: missense mutations suppress gene silencing associated with position effect variegation. *Genetics* **154**, 657-668 (2000).

134 Mis, J., Ner, S. S. & Grigliatti, T. A. Identification of three histone methyltransferases in Drosophila: dG9a is a suppressor of PEV and is required for gene silencing. *Mol Genet Genomics* **275**, 513-526, doi:10.1007/s00438-006-0116-x (2006).

135 Baksa, K. *et al.* Mutations in the protein phosphatase 1 gene at 87B can differentially affect suppression of position-effect variegation and mitosis in Drosophila melanogaster. *Genetics* **135**, 117-125 (1993).

136 Csink, A. K., Linsk, R. & Birchler, J. A. The Lighten up (Lip) gene of Drosophila melanogaster, a modifier of retroelement expression, position effect variegation and white locus insertion alleles. *Genetics* **138**, 153-163 (1994).

137 Gracheva, E., Dus, M. & Elgin, S. C. Drosophila RISC component VIG and its homolog Vig2 impact heterochromatin formation. *PLoS One* **4**, e6182, doi:10.1371/journal.pone.0006182 (2009).

138 Tan, W., Schauder, C., Naryshkina, T., Minakhina, S. & Steward, R. Zfrp8 forms a complex with fragile-X mental retardation protein and regulates its localization and function. *Dev Biol* **410**, 202-212, doi:10.1016/j.ydbio.2015.12.008 (2016).

139 Emelyanov, A. V., Konev, A. Y., Vershilova, E. & Fyodorov, D. V. Protein complex of Drosophila ATRX/XNP and HP1a is required for the formation of pericentric beta-heterochromatin in vivo. *J Biol Chem* **285**, 15027-15037, doi:10.1074/jbc.M109.064790 (2010).

140 James, T. C. & Elgin, S. C. Identification of a nonhistone chromosomal protein associated with heterochromatin in Drosophila melanogaster and its gene. *Mol Cell Biol* **6**, 3862-3872, doi:10.1128/mcb.6.11.3862 (1986).

141 Hari, K. L., Cook, K. R. & Karpen, G. H. The Drosophila Su(var)2-10 locus regulates chromosome structure and function and encodes a member of the PIAS protein family. *Genes Dev* **15**, 1334-1348, doi:10.1101/gad.877901 (2001).

142 Casas-Vila, N. *et al.* The developmental proteome of Drosophila melanogaster. *Genome Res* **27**, 1273-1285, doi:10.1101/gr.213694.116 (2017).

143 Hari, K. *Su(var)2-10 is an Essential Component of Interphase Chromosome Organization* Doctor of Philosophy in Biology thesis, University of California, San Diego, (2001).

144 Kallenbach, L. *Role of the SUMO Ligase PIAS in Chromatin Organization* PhD thesis, Albert-Ludwigs-Universität, (2011).

145 Wang, R. *et al.* The conserved ancient role of chordate PIAS as a multilevel repressor of the NF-κB pathway. *Sci Rep* **7**, 17063, doi:10.1038/s41598-017-16624-7 (2017).

146 Aravind, L. & Koonin, E. V. SAP - a putative DNA-binding motif involved in chromosomal organization. *Trends Biochem Sci* **25**, 112-114 (2000).

147 Sachdev, S. *et al.* PIASy, a nuclear matrix-associated SUMO E3 ligase, represses LEF1 activity by sequestration into nuclear bodies. *Genes Dev* **15**, 3088-3103 (2001).

148 Duval, D., Duval, G., Kedinger, C., Poch, O. & Boeuf, H. The 'PINIT' motif, of a newly identified conserved domain of the PIAS protein family, is essential for nuclear retention of PIAS3L. *FEBS Lett* **554**, 111-118 (2003).
149 Yunus, A. A. & Lima, C. D. Structure of the Siz/PIAS SUMO E3 ligase Siz1 and determinants required for SUMO modification of PCNA. *Mol Cell* **35**, 669-682, doi:10.1016/j.molcel.2009.07.013 (2009).
150 Schmidt, D. & Müller, S. Members of the PIAS family act as SUMO ligases for c-Jun and p53 and repress p53 activity. *Proc Natl Acad Sci U S A* **99**, 2872-2877, doi:10.1073/pnas.052559499 (2002).
151 Kaur, K., Park, H., Pandey, N., Azuma, Y. & De Guzman, R. N. Identification of a new small ubiquitin-like modifier (SUMO)-interacting motif in the E3 ligase PIASy. *J Biol Chem* **292**, 10230-10238, doi:10.1074/jbc.M117.789982 (2017).
152 Betz, A., Lampen, N., Martinek, S., Young, M. W. & Darnell, J. E. A Drosophila PIAS homologue negatively regulates stat92E. *Proc Natl Acad Sci U S A* **98**, 9563-9568, doi:10.1073/pnas.171302098 (2001).
153 Kalamarz, M. E., Paddibhatla, I., Nadar, C. & Govind, S. Sumoylation is tumor-suppressive and confers proliferative quiescence to hematopoietic progenitors in Drosophila melanogaster larvae. *Biol Open* **1**, 161-172, doi:10.1242/bio.2012043 (2012).
154 Driscoll, J. J. *et al.* The sumoylation pathway is dysregulated in multiple myeloma and is associated with adverse patient outcome. *Blood* **115**, 2827-2834, doi:10.1182/blood-2009-03-211045 (2010).
155 Chung, C. D. *et al.* Specific inhibition of Stat3 signal transduction by PIAS3. *Science* **278**, 1803-1805 (1997).
156 Liu, B. *et al.* Inhibition of Stat1-mediated gene activation by PIAS1. *Proc Natl Acad Sci U S A* **95**, 10626-10631 (1998).
157 Liu, B. *et al.* Negative regulation of NF-kappaB signaling by PIAS1. *Mol Cell Biol* **25**, 1113-1123, doi:10.1128/MCB.25.3.1113-1123.2005 (2005).
158 Constanzo, J. D. *et al.* Pias1 is essential for erythroid and vascular development in the mouse embryo. *Dev Biol* **415**, 98-110, doi:10.1016/j.ydbio.2016.04.013 (2016).
159 Wong, K. A. *et al.* Protein inhibitor of activated STAT Y (PIASy) and a splice variant lacking exon 6 enhance sumoylation but are not essential for embryogenesis and adult life. *Mol Cell Biol* **24**, 5577-5586, doi:10.1128/MCB.24.12.5577-5586.2004 (2004).
160 Roth, W. *et al.* PIASy-deficient mice display modest defects in IFN and Wnt signaling. *J Immunol* **173**, 6189-6199 (2004).
161 Liu, B. *et al.* PIAS1 selectively inhibits interferon-inducible genes and is important in innate immunity. *Nat Immunol* **5**, 891-898, doi:10.1038/ni1104 (2004).
162 Santti, H. *et al.* Disruption of the murine PIASx gene results in reduced testis weight. *J Mol Endocrinol* **34**, 645-654, doi:10.1677/jme.1.01666 (2005).
163 Maria, N. *et al.* (https://doi.org/10.1101/533091, pre-print 2019).
164 Muerdter, F. *et al.* A genome-wide RNAi screen draws a genetic framework for transposon control and primary piRNA biogenesis in Drosophila. *Mol Cell* **50**, 736-748, doi:10.1016/j.molcel.2013.04.006 (2013).
165 Kumar, D. *et al.* NMR-derived solution structure of SUMO from Drosophila melanogaster (dSmt3). *Proteins* **75**, 1046-1050, doi:10.1002/prot.22389 (2009).

166 *SUMO Regulation of Cellular Processes*. (Caister Academic Pres, 2017).
167 Mannen, H., Tseng, H. M., Cho, C. L. & Li, S. S. Cloning and expression of human homolog HSMT3 to yeast SMT3 suppressor of MIF2 mutations in a centromere protein gene. *Biochem Biophys Res Commun* **222**, 178-180, doi:10.1006/bbrc.1996.0717 (1996).
168 Meluh, P. B. & Koshland, D. Evidence that the MIF2 gene of Saccharomyces cerevisiae encodes a centromere protein with homology to the mammalian centromere protein CENP-C. *Mol Biol Cell* **6**, 793-807 (1995).
169 Hendriks, I. A. *et al.* Site-specific mapping of the human SUMO proteome reveals co-modification with phosphorylation. *Nat Struct Mol Biol* **24**, 325-336, doi:10.1038/nsmb.3366 (2017).
170 Chung, T. L. *et al.* In vitro modification of human centromere protein CENP-C fragments by small ubiquitin-like modifier (SUMO) protein: definitive identification of the modification sites by tandem mass spectrometry analysis of the isopeptides. *J Biol Chem* **279**, 39653-39662, doi:10.1074/jbc.M405637200 (2004).
171 Nie, M., Xie, Y., Loo, J. A. & Courey, A. J. Genetic and proteomic evidence for roles of Drosophila SUMO in cell cycle control, Ras signaling, and early pattern formation. *PLoS One* **4**, e5905, doi:10.1371/journal.pone.0005905 (2009).
172 Kroetz, M. B. SUMO: a ubiquitin-like protein modifier. *Yale J Biol Med* **78**, 197-201 (2005).
173 Ulrich, H. D. The SUMO system: an overview. *Methods Mol Biol* **497**, 3-16, doi:10.1007/978-1-59745-566-4_1 (2009).
174 Sasidharan, K., Soga, T., Tomita, M. & Murray, D. B. A yeast metabolite extraction protocol optimised for time-series analyses. *PLoS One* **7**, e44283, doi:10.1371/journal.pone.0044283 (2012).
175 Rodriguez, M. S., Dargemont, C. & Hay, R. T. SUMO-1 conjugation in vivo requires both a consensus modification motif and nuclear targeting. *J Biol Chem* **276**, 12654-12659, doi:10.1074/jbc.M009476200 (2001).
176 Hendriks, I. A. *et al.* Uncovering global SUMOylation signaling networks in a site-specific manner. *Nat Struct Mol Biol* **21**, 927-936, doi:10.1038/nsmb.2890 (2014).
177 Beauclair, G., Bridier-Nahmias, A., Zagury, J. F., Saïb, A. & Zamborlini, A. JASSA: a comprehensive tool for prediction of SUMOylation sites and SIMs. *Bioinformatics* **31**, 3483-3491, doi:10.1093/bioinformatics/btv403 (2015).
178 Zhao, Q. *et al.* GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs. *Nucleic Acids Res* **42**, W325-330, doi:10.1093/nar/gku383 (2014).
179 Pichler, A., Fatouros, C., Lee, H. & Eisenhardt, N. SUMO conjugation - a mechanistic view. *Biomol Concepts* **8**, 13-36, doi:10.1515/bmc-2016-0030 (2017).
180 Song, J., Zhang, Z., Hu, W. & Chen, Y. Small ubiquitin-like modifier (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation. *J Biol Chem* **280**, 40122-40129, doi:10.1074/jbc.M507059200 (2005).
181 Kerscher, O. SUMO junction-what's your function? New insights through SUMO-interacting motifs. *EMBO Rep* **8**, 550-555, doi:10.1038/sj.embor.7400980 (2007).

182 Hecker, C. M., Rabiller, M., Haglund, K., Bayer, P. & Dikic, I. Specification of SUMO1- and SUMO2-interacting motifs. *J Biol Chem* **281**, 16117-16127, doi:10.1074/jbc.M512757200 (2006).
183 Hietakangas, V. *et al.* PDSM, a motif for phosphorylation-dependent SUMO modification. *Proc Natl Acad Sci U S A* **103**, 45-50, doi:10.1073/pnas.0503698102 (2006).
184 Perry, J. J., Tainer, J. A. & Boddy, M. N. A SIM-ultaneous role for SUMO and ubiquitin. *Trends Biochem Sci* **33**, 201-208, doi:10.1016/j.tibs.2008.02.001 (2008).
185 Mario, G.-D. *SUMO Tasks in Chromatin Remodelling*, 2012).
186 Hannoun, Z., Greenhough, S., Jaffray, E., Hay, R. T. & Hay, D. C. Post-translational modification by SUMO. *Toxicology* **278**, 288-293, doi:10.1016/j.tox.2010.07.013 (2010).
187 Johnson, E. S. Protein modification by SUMO. *Annu Rev Biochem* **73**, 355-382, doi:10.1146/annurev.biochem.73.011303.074118 (2004).
188 Wang, L. *et al.* SUMO2 is essential while SUMO3 is dispensable for mouse embryonic development. *EMBO Rep* **15**, 878-885, doi:10.15252/embr.201438534 (2014).
189 Zhang, F. P. *et al.* Sumo-1 function is dispensable in normal mouse development. *Mol Cell Biol* **28**, 5381-5390, doi:10.1128/MCB.00651-08 (2008).
190 Nacerddine, K. *et al.* The SUMO pathway is essential for nuclear integrity and chromosome segregation in mice. *Dev Cell* **9**, 769-779, doi:10.1016/j.devcel.2005.10.007 (2005).
191 Huang, L., Ohsako, S. & Tanda, S. The lesswright mutation activates Rel-related proteins, leading to overproduction of larval hemocytes in Drosophila melanogaster. *Dev Biol* **280**, 407-420, doi:10.1016/j.ydbio.2005.02.006 (2005).
192 Kanakousaki, K. & Gibson, M. C. A differential requirement for SUMOylation in proliferating and non-proliferating cells during Drosophila development. *Development* **139**, 2751-2762, doi:10.1242/dev.082974 (2012).
193 Smith, M., Turki-Judeh, W. & Courey, A. J. SUMOylation in Drosophila Development. *Biomolecules* **2**, 331-349, doi:10.3390/biom2030331 (2012).
194 Drabikowski, K. *et al.* Comprehensive list of SUMO targets in Caenorhabditis elegans and its implication for evolutionary conservation of SUMO signaling. *Sci Rep* **8**, 1139, doi:10.1038/s41598-018-19424-9 (2018).
195 Shin, J. A. *et al.* SUMO modification is involved in the maintenance of heterochromatin stability in fission yeast. *Mol Cell* **19**, 817-828, doi:10.1016/j.molcel.2005.08.021 (2005).
196 Ihara, M., Stein, P. & Schultz, R. M. UBE2I (UBC9), a SUMO-conjugating enzyme, localizes to nuclear speckles and stimulates transcription in mouse oocytes. *Biol Reprod* **79**, 906-913, doi:10.1095/biolreprod.108.070474 (2008).
197 Richard, P., Vethantham, V. & Manley, J. L. Roles of Sumoylation in mRNA Processing and Metabolism. *Adv Exp Med Biol* **963**, 15-33, doi:10.1007/978-3-319-50044-7_2 (2017).
198 Pozzi, B. *et al.* SUMO conjugation to spliceosomal proteins is required for efficient pre-mRNA splicing. *Nucleic Acids Res* **45**, 6729-6745, doi:10.1093/nar/gkx213 (2017).

199 Hendriks, I. A. & Vertegaal, A. C. A comprehensive compilation of SUMO proteomics. *Nat Rev Mol Cell Biol* **17**, 581-595, doi:10.1038/nrm.2016.81 (2016).

200 Hendriks, I. A. *et al.* Site-specific characterization of endogenous SUMOylation across species and organs. *Nat Commun* **9**, 2456, doi:10.1038/s41467-018-04957-4 (2018).

201 McManus, F. P. *et al.* Quantitative SUMO proteomics reveals the modulation of several PML nuclear body associated proteins and an anti-senescence function of UBC9. *Sci Rep* **8**, 7754, doi:10.1038/s41598-018-25150-z (2018).

202 Uzoma, I. *et al.* Global Identification of Small Ubiquitin-related Modifier (SUMO) Substrates Reveals Crosstalk between SUMOylation and Phosphorylation Promotes Cell Migration. *Mol Cell Proteomics* **17**, 871-888, doi:10.1074/mcp.RA117.000014 (2018).

203 Cai, L. *et al.* Proteome-wide Mapping of Endogenous SUMOylation Sites in Mouse Testis. *Mol Cell Proteomics* **16**, 717-727, doi:10.1074/mcp.M116.062125 (2017).

204 Yang, W. & Paschen, W. SUMO proteomics to decipher the SUMO-modified proteome regulated by various diseases. *Proteomics* **15**, 1181-1191, doi:10.1002/pmic.201400298 (2015).

205 Tirard, M. *et al.* In vivo localization and identification of SUMOylated proteins in the brain of His6-HA-SUMO1 knock-in mice. *Proc Natl Acad Sci U S A* **109**, 21122-21127, doi:10.1073/pnas.1215366110 (2012).

206 Kaminsky, R. *et al.* SUMO regulates the assembly and function of a cytoplasmic intermediate filament protein in C. elegans. *Dev Cell* **17**, 724-735, doi:10.1016/j.devcel.2009.10.005 (2009).

207 Handu, M. *et al.* SUMO-Enriched Proteome for Drosophila Innate Immune Response. *G3 (Bethesda)* **5**, 2137-2154, doi:10.1534/g3.115.020958 (2015).

208 Denison, C. *et al.* A proteomic strategy for gaining insights into protein sumoylation in yeast. *Mol Cell Proteomics* **4**, 246-254, doi:10.1074/mcp.M400154-MCP200 (2005).

209 Yang, W., Sheng, H. & Wang, H. Targeting the SUMO pathway for neuroprotection in brain ischaemia. *Stroke Vasc Neurol* **1**, 101-107, doi:10.1136/svn-2016-000031 (2016).

210 Tammsalu, T. *et al.* Proteome-wide identification of SUMO modification sites by mass spectrometry. *Nat Protoc* **10**, 1374-1388, doi:10.1038/nprot.2015.095 (2015).

211 Rosonina, E., Akhter, A., Dou, Y., Babu, J. & Sri Theivakadadcham, V. S. Regulation of transcription factors by sumoylation. *Transcription* **8**, 220-231, doi:10.1080/21541264.2017.1311829 (2017).

212 Rosonina, E. A conserved role for transcription factor sumoylation in binding-site selection. *Curr Genet*, doi:10.1007/s00294-019-00992-w (2019).

213 Cossec, J. C. *et al.* SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing Distinct Chromatin States. *Cell Stem Cell* **23**, 742-757.e748, doi:10.1016/j.stem.2018.10.001 (2018).

214 Tsurumi, A. *et al.* STAT is an essential activator of the zygotic genome in the early Drosophila embryo. *PLoS Genet* **7**, e1002086, doi:10.1371/journal.pgen.1002086 (2011).

215 Grönholm, J., Ungureanu, D., Vanhatupa, S., Rämet, M. & Silvennoinen, O. Sumoylation of Drosophila transcription factor STAT92E. *J Innate Immun* **2**, 618-624, doi:10.1159/000318676 (2010).

216 Vethantham, V., Rao, N. & Manley, J. L. Sumoylation regulates multiple aspects of mammalian poly(A) polymerase function. *Genes Dev* **22**, 499-511, doi:10.1101/gad.1628208 (2008).
217 Vethantham, V., Rao, N. & Manley, J. L. Sumoylation modulates the assembly and activity of the pre-mRNA 3' processing complex. *Mol Cell Biol* **27**, 8848-8858, doi:10.1128/MCB.01186-07 (2007).
218 Vassileva, M. T. & Matunis, M. J. SUMO modification of heterogeneous nuclear ribonucleoproteins. *Mol Cell Biol* **24**, 3623-3632, doi:10.1128/mcb.24.9.3623-3632.2004 (2004).
219 Du, Y. *et al.* SUMOylation of the m6A-RNA methyltransferase METTL3 modulates its function. *Nucleic Acids Res* **46**, 5195-5208, doi:10.1093/nar/gky156 (2018).
220 Rappsilber, J., Ryder, U., Lamond, A. I. & Mann, M. Large-scale proteomic analysis of the human spliceosome. *Genome Res* **12**, 1231-1245, doi:10.1101/gr.473902 (2002).
221 Tan, J. A. *et al.* Protein inhibitors of activated STAT resemble scaffold attachment factors and function as interacting nuclear receptor coregulators. *J Biol Chem* **277**, 16993-17001, doi:10.1074/jbc.M109217200 (2002).
222 Lamond, A. I. & Spector, D. L. Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Biol* **4**, 605-612, doi:10.1038/nrm1172 (2003).
223 Jentsch, S. & Psakhye, I. Control of nuclear activities by substrate-selective and protein-group SUMOylation. *Annu Rev Genet* **47**, 167-186, doi:10.1146/annurev-genet-111212-133453 (2013).
224 Hendriks, I. A. & Vertegaal, A. C. SUMO in the DNA damage response. *Oncotarget* **6**, 15734-15735, doi:10.18632/oncotarget.4605 (2015).
225 Psakhye, I. & Jentsch, S. Protein group modification and synergy in the SUMO pathway as exemplified in DNA repair. *Cell* **151**, 807-820, doi:10.1016/j.cell.2012.10.021 (2012).
226 Capelson, M. & Corces, V. G. SUMO conjugation attenuates the activity of the gypsy chromatin insulator. *EMBO J* **25**, 1906-1914, doi:10.1038/sj.emboj.7601068 (2006).
227 Capelson, M. & Corces, V. G. The ubiquitin ligase dTopors directs the nuclear organization of a chromatin insulator. *Mol Cell* **20**, 105-116, doi:10.1016/j.molcel.2005.08.031 (2005).
228 Pungaliya, P. *et al.* TOPORS functions as a SUMO-1 E3 ligase for chromatin-modifying proteins. *J Proteome Res* **6**, 3918-3923, doi:10.1021/pr0703674 (2007).
229 Secombe, J. & Parkhurst, S. M. Drosophila Topors is a RING finger-containing protein that functions as a ubiquitin-protein isopeptide ligase for the hairy basic helix-loop-helix repressor protein. *J Biol Chem* **279**, 17126-17133, doi:10.1074/jbc.M310097200 (2004).
230 Matsui, M. *et al.* Nuclear structure and chromosome segregation in Drosophila male meiosis depend on the ubiquitin ligase dTopors. *Genetics* **189**, 779-793, doi:10.1534/genetics.111.133819 (2011).
231 Shiio, Y. & Eisenman, R. N. Histone sumoylation is associated with transcriptional repression. *Proc Natl Acad Sci U S A* **100**, 13225-13230, doi:10.1073/pnas.1735528100 (2003).
232 Nathan, D. *et al.* Histone sumoylation is a negative regulator in Saccharomyces cerevisiae and shows dynamic interplay with positive-acting

histone modifications. *Genes Dev* **20**, 966-976, doi:10.1101/gad.1404206 (2006).
233 Galisson, F. *et al.* A novel proteomics approach to identify SUMOylated proteins and their modification sites in human cells. *Mol Cell Proteomics* **10**, M110.004796, doi:10.1074/mcp.M110.004796 (2011).
234 Dhall, A. *et al.* Sumoylated human histone H4 prevents chromatin compaction by inhibiting long-range internucleosomal interactions. *J Biol Chem* **289**, 33827-33837, doi:10.1074/jbc.M114.591644 (2014).
235 Dhall, A., Weller, C. E., Chu, A., Shelton, P. M. M. & Chatterjee, C. Chemically Sumoylated Histone H4 Stimulates Intranucleosomal Demethylation by the LSD1-CoREST Complex. *ACS Chem Biol* **12**, 2275-2280, doi:10.1021/acschembio.7b00716 (2017).
236 Sáez, J. E., Arredondo, C., Rivera, C. & Andrés, M. E. PIASγ controls stability and facilitates SUMO-2 conjugation to CoREST family of transcriptional co-repressors. *Biochem J* **475**, 1441-1454, doi:10.1042/BCJ20170983 (2018).
237 David, G., Neptune, M. A. & DePinho, R. A. SUMO-1 modification of histone deacetylase 1 (HDAC1) modulates its biological activities. *J Biol Chem* **277**, 23658-23663, doi:10.1074/jbc.M203690200 (2002).
238 Tao, C. C., Hsu, W. L., Ma, Y. L., Cheng, S. J. & Lee, E. H. Epigenetic regulation of HDAC1 SUMOylation as an endogenous neuroprotection against Aβ toxicity in a mouse model of Alzheimer's disease. *Cell Death Differ* **24**, 597-614, doi:10.1038/cdd.2016.161 (2017).
239 Gocke, C. B. & Yu, H. ZNF198 stabilizes the LSD1-CoREST-HDAC1 complex on chromatin through its MYM-type zinc fingers. *PLoS One* **3**, e3255, doi:10.1371/journal.pone.0003255 (2008).
240 David, G., Neptune, M. A. & DePinho, R. A. SUMO-1 modification of histone deacetylase 1 (HDAC1) modulates its biological activities. *J Biol Chem* **277**, 23658-23663, doi:10.1074/jbc.M203690200 (2002).
241 Lyst, M. J., Nan, X. & Stancheva, I. Regulation of MBD1-mediated transcriptional repression by SUMO and PIAS proteins. *EMBO J* **25**, 5317-5328, doi:10.1038/sj.emboj.7601404 (2006).
242 Uchimura, Y. *et al.* Involvement of SUMO modification in MBD1- and MCAF1-mediated heterochromatin formation. *J Biol Chem* **281**, 23180-23190, doi:10.1074/jbc.M602280200 (2006).
243 Josa-Prado, F., Henley, J. M. & Wilkinson, K. A. SUMOylation of Argonaute-2 regulates RNA interference activity. *Biochem Biophys Res Commun* **464**, 1066-1071, doi:10.1016/j.bbrc.2015.07.073 (2015).
244 Maison, C. *et al.* SUMOylation promotes de novo targeting of HP1α to pericentric heterochromatin. *Nat Genet* **43**, 220-227, doi:10.1038/ng.765 (2011).
245 Muchardt, C. *et al.* Coordinated methyl and RNA binding is required for heterochromatin localization of mammalian HP1alpha. *EMBO Rep* **3**, 975-981, doi:10.1093/embo-reports/kvf194 (2002).
246 Müller, M. M., Fierz, B., Bittova, L., Liszczak, G. & Muir, T. W. A two-state activation mechanism controls the histone methyltransferase Suv39h1. *Nat Chem Biol* **12**, 188-193, doi:10.1038/nchembio.2008 (2016).
247 Maison, C., Bailly, D., Quivy, J. P. & Almouzni, G. The methyltransferase Suv39h1 links the SUMO pathway to HP1α marking at pericentric heterochromatin. *Nat Commun* **7**, 12224, doi:10.1038/ncomms12224 (2016).

248 Kang, Y. K. SETDB1 in Early Embryos and Embryonic Stem Cells. *Curr Issues Mol Biol* **17**, 1-10 (2015).
249 Le Douarin, B. *et al.* A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic control of transcription by nuclear receptors. *EMBO J* **15**, 6701-6715 (1996).
250 Jang, S. M. *et al.* KAP1 facilitates reinstatement of heterochromatin after DNA replication. *Nucleic Acids Res* **46**, 8788-8802, doi:10.1093/nar/gky580 (2018).
251 Peng, J. & Wysocka, J. It takes a PHD to SUMO. *Trends Biochem Sci* **33**, 191-194, doi:10.1016/j.tibs.2008.02.003 (2008).
252 Gibbons, R. J., Picketts, D. J., Villard, L. & Higgs, D. R. Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). *Cell* **80**, 837-845 (1995).
253 Picketts, D. J. *et al.* ATRX encodes a novel member of the SNF2 family of proteins: mutations point to a common mechanism underlying the ATR-X syndrome. *Hum Mol Genet* **5**, 1899-1907, doi:10.1093/hmg/5.12.1899 (1996).
254 Argentaro, A. *et al.* Structural consequences of disease-causing mutations in the ATRX-DNMT3-DNMT3L (ADD) domain of the chromatin-associated protein ATRX. *Proc Natl Acad Sci U S A* **104**, 11939-11944, doi:10.1073/pnas.0704057104 (2007).
255 Iwase, S. *et al.* ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental-retardation syndrome. *Nat Struct Mol Biol* **18**, 769-776, doi:10.1038/nsmb.2062 (2011).
256 Otani, J. *et al.* Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep* **10**, 1235-1241, doi:10.1038/embor.2009.218 (2009).
257 Eustermann, S. *et al.* Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin. *Nat Struct Mol Biol* **18**, 777-782, doi:10.1038/nsmb.2070 (2011).
258 Dhayalan, A. *et al.* The ATRX-ADD domain binds to H3 tail peptides and reads the combined methylation state of K4 and K9. *Hum Mol Genet* **20**, 2195-2203, doi:10.1093/hmg/ddr107 (2011).
259 Law, M. J. *et al.* ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell* **143**, 367-378, doi:10.1016/j.cell.2010.09.023 (2010).
260 Cree, S. L. *et al.* DNA G-quadruplexes show strong interaction with DNA methyltransferases in vitro. *FEBS Lett* **590**, 2870-2883, doi:10.1002/1873-3468.12331 (2016).
261 Lechner, M. S., Schultz, D. C., Negorev, D., Maul, G. G. & Rauscher, F. J. The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain. *Biochem Biophys Res Commun* **331**, 929-937, doi:10.1016/j.bbrc.2005.04.016 (2005).
262 Kernohan, K. D. *et al.* ATRX partners with cohesin and MeCP2 and contributes to developmental silencing of imprinted genes in the brain. *Dev Cell* **18**, 191-202, doi:10.1016/j.devcel.2009.12.017 (2010).
263 Garrick, D. *et al.* Loss of Atrx affects trophoblast development and the pattern of X-inactivation in extraembryonic tissues. *PLoS Genet* **2**, e58, doi:10.1371/journal.pgen.0020058 (2006).

264 Abidi, F. *et al.* Carpenter-Waziri syndrome results from a mutation in XNP. *Am J Med Genet* **85**, 249-251 (1999).
265 Lossi, A. M. *et al.* Mutation of the XNP/ATR-X gene in a family with severe mental retardation, spastic paraplegia and skewed pattern of X inactivation: demonstration that the mutation is involved in the inactivation bias. *Am J Hum Genet* **65**, 558-562, doi:10.1086/302499 (1999).
266 Villard, L. *et al.* XNP mutation in a large family with Juberg-Marsidi syndrome. *Nat Genet* **12**, 359-360, doi:10.1038/ng0496-359 (1996).
267 Wada, T., Sugie, H., Fukushima, Y. & Saitoh, S. Non-skewed X-inactivation may cause mental retardation in a female carrier of X-linked alpha-thalassemia/mental retardation syndrome (ATR-X): X-inactivation study of nine female carriers of ATR-X. *Am J Med Genet A* **138**, 18-20, doi:10.1002/ajmg.a.30901 (2005).
268 Wada, T. *et al.* Neuroradiologic features in X-linked α-thalassemia/mental retardation syndrome. *AJNR Am J Neuroradiol* **34**, 2034-2038, doi:10.3174/ajnr.A3560 (2013).
269 Levy, M. A., Kernohan, K. D., Jiang, Y. & Bérubé, N. G. ATRX promotes gene expression by facilitating transcriptional elongation through guanine-rich coding regions. *Hum Mol Genet* **24**, 1824-1835, doi:10.1093/hmg/ddu596 (2015).
270 Wang, Y. *et al.* G-quadruplex DNA drives genomic instability and represents a targetable molecular abnormality in ATRX-deficient malignant glioma. *Nat Commun* **10**, 943, doi:10.1038/s41467-019-08905-8 (2019).
271 Nguyen, D. T. *et al.* The chromatin remodelling factor ATRX suppresses R-loops in transcribed telomeric repeats. *EMBO Rep* **18**, 914-928, doi:10.15252/embr.201643078 (2017).
272 McDowell, T. L. *et al.* Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. *Proc Natl Acad Sci U S A* **96**, 13983-13988, doi:10.1073/pnas.96.24.13983 (1999).
273 Ishov, A. M., Vladimirova, O. V. & Maul, G. G. Heterochromatin and ND10 are cell-cycle regulated and phosphorylation-dependent alternate nuclear sites of the transcription repressor Daxx and SWI/SNF protein ATRX. *J Cell Sci* **117**, 3807-3820, doi:10.1242/jcs.01230 (2004).
274 Sadic, D. *et al.* Atrx promotes heterochromatin formation at retrotransposons. *EMBO Rep* **16**, 836-850, doi:10.15252/embr.201439937 (2015).
275 He, Q. *et al.* The Daxx/Atrx Complex Protects Tandem Repetitive Elements during DNA Hypomethylation by Promoting H3K9 Trimethylation. *Cell Stem Cell* **17**, 273-286, doi:10.1016/j.stem.2015.07.022 (2015).
276 De La Fuente, R., Viveiros, M. M., Wigglesworth, K. & Eppig, J. J. ATRX, a member of the SNF2 family of helicase/ATPases, is required for chromosome alignment and meiotic spindle organization in metaphase II stage mouse oocytes. *Dev Biol* **272**, 1-14, doi:10.1016/j.ydbio.2003.12.012 (2004).
277 Lovejoy, C. A. *et al.* Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of telomeres pathway. *PLoS Genet* **8**, e1002772, doi:10.1371/journal.pgen.1002772 (2012).

278 Schwartz, B. E. & Ahmad, K. Transcriptional activation triggers deposition and removal of the histone variant H3.3. *Genes Dev* **19**, 804-814, doi:10.1101/gad.1259805 (2005).
279 Kaplan, N. *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**, 362-366, doi:10.1038/nature07667 (2009).
280 Valle-García, D. *et al.* ATRX binds to atypical chromatin domains at the 3' exons of zinc finger genes to preserve H3K9me3 enrichment. *Epigenetics* **11**, 398-414, doi:10.1080/15592294.2016.1169351 (2016).
281 Gauchier, M. *et al.* SETDB1-dependent heterochromatin stimulates alternative lengthening of telomeres. *Sci Adv* **5**, eaav3673, doi:10.1126/sciadv.aav3673 (2019).
282 Tardat, M. & Déjardin, J. Telomere chromatin establishment and its maintenance during mammalian development. *Chromosoma* **127**, 3-18, doi:10.1007/s00412-017-0656-3 (2018).
283 Gibbons, R. J. *et al.* Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. *Nat Genet* **24**, 368-371, doi:10.1038/74191 (2000).
284 Kernohan, K. D., Vernimmen, D., Gloor, G. B. & Bérubé, N. G. Analysis of neonatal brain lacking ATRX or MeCP2 reveals changes in nucleosome density, CTCF binding and chromatin looping. *Nucleic Acids Res* **42**, 8356-8368, doi:10.1093/nar/gku564 (2014).
285 Probst, A. V. *et al.* A strand-specific burst in transcription of pericentric satellites is required for chromocenter formation and early mouse development. *Dev Cell* **19**, 625-638, doi:10.1016/j.devcel.2010.09.002 (2010).
286 Chu, H. P. *et al.* TERRA RNA Antagonizes ATRX and Protects Telomeres. *Cell* **170**, 86-101.e116, doi:10.1016/j.cell.2017.06.017 (2017).
287 Hirschi, A., Martin, W. J., Luka, Z., Loukachevitch, L. V. & Reiter, N. J. G-quadruplex RNA binding and recognition by the lysine-specific histone demethylase-1 enzyme. *RNA* **22**, 1250-1260, doi:10.1261/rna.057265.116 (2016).
288 Arora, R. *et al.* RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nat Commun* **5**, 5220, doi:10.1038/ncomms6220 (2014).
289 Schneiderman, J. I., Sakai, A., Goldstein, S. & Ahmad, K. The XNP remodeler targets dynamic chromatin in Drosophila. *Proc Natl Acad Sci U S A* **106**, 14472-14477, doi:10.1073/pnas.0905816106 (2009).
290 Duc, C. *et al.* Arabidopsis ATRX Modulates H3.3 Occupancy and Fine-Tunes Gene Expression. *Plant Cell* **29**, 1773-1793, doi:10.1105/tpc.16.00877 (2017).
291 Cardoso, C. *et al.* XNP-1/ATR-X acts with RB, HP1 and the NuRD complex during larval development in C. elegans. *Dev Biol* **278**, 49-59, doi:10.1016/j.ydbio.2004.10.014 (2005).
292 Xue, Y. *et al.* The ATRX syndrome protein forms a chromatin-remodeling complex with Daxx and localizes in promyelocytic leukemia nuclear bodies. *Proc Natl Acad Sci U S A* **100**, 10635-10640, doi:10.1073/pnas.1937626100 (2003).
293 Emelyanov, A. V., Konev, A. Y., Vershilova, E. & Fyodorov, D. V. Protein complex of Drosophila ATRX/XNP and HP1a is required for the formation of

pericentric beta-heterochromatin in vivo. *J Biol Chem* **285**, 15027-15037, doi:10.1074/jbc.M109.064790 (2010).

294 Bassett, A. R., Cooper, S. E., Ragab, A. & Travers, A. A. The chromatin remodelling factor dATRX is involved in heterochromatin formation. *PLoS One* **3**, e2099, doi:10.1371/journal.pone.0002099 (2008).

295 Chavez, J. *et al.* dAdd1 and dXNP prevent genome instability by maintaining HP1a localization at Drosophila telomeres. *Chromosoma* **126**, 697-712, doi:10.1007/s00412-017-0634-9 (2017).

296 López-Falcón, B. *et al.* Characterization of the Drosophila group ortholog to the amino-terminus of the alpha-thalassemia and mental retardation X-Linked (ATRX) vertebrate protein. *PLoS One* **9**, e113182, doi:10.1371/journal.pone.0113182 (2014).

297 Jagannathan, M., Warsinger-Pepe, N., Watase, G. J. & Yamashita, Y. M. Comparative Analysis of Satellite DNA in the. *G3 (Bethesda)* **7**, 693-704, doi:10.1534/g3.116.035352 (2017).

298 Jagannathan, M., Cummings, R. & Yamashita, Y. M. A conserved function for pericentromeric satellite DNA. *Elife* **7**, doi:10.7554/eLife.34122 (2018).

299 Alekseyenko, A. A. *et al.* Heterochromatin-associated interactions of Drosophila HP1a with dADD1, HIPP1, and repetitive RNAs. *Genes Dev* **28**, 1445-1460, doi:10.1101/gad.241950.114 (2014).

300 Beckstead, R. B. *et al.* Bonus, a Drosophila TIF1 homolog, is a chromatin-associated protein that acts as a modifier of position-effect variegation. *Genetics* **169**, 783-794, doi:10.1534/genetics.104.037085 (2005).

301 Le Douarin, B., You, J., Nielsen, A. L., Chambon, P. & Losson, R. TIF1alpha: a possible link between KRAB zinc finger proteins and nuclear receptors. *J Steroid Biochem Mol Biol* **65**, 43-50 (1998).

302 Nielsen, A. L. *et al.* Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. *EMBO J* **18**, 6385-6395, doi:10.1093/emboj/18.22.6385 (1999).

303 Swenson, J. M., Colmenares, S. U., Strom, A. R., Costes, S. V. & Karpen, G. H. The composition and organization of Drosophila heterochromatin are heterogeneous and dynamic. *Elife* **5**, doi:10.7554/eLife.16096 (2016).

304 Steuernagel, Y. *Characterization of the dPIAS-dATRX Interaction* Master of Biological Sciences thesis, The University of Constance, (2012).

305 Golebiowski, F. *et al.* System-wide changes to SUMO modifications in response to heat shock. *Sci Signal* **2**, ra24, doi:10.1126/scisignal.2000282 (2009).

306 Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **75**, 663-670 (2003).

307 Lyne, R. *et al.* FlyMine: an integrated database for Drosophila and Anopheles genomics. *Genome Biol* **8**, R129, doi:10.1186/gb-2007-8-7-r129 (2007).

308 Gangadharan, Y. D. *THE ROLE OF DROSOPHILA SUMO E3 LIGASE dPIAS IN CHROMOSOME ORGANIZATION* PhD thesis, der Albert-Ludwigs-Universität Freiburg im Breisgau, (2010).

309 Henikoff, S., Henikoff, J. G., Sakai, A., Loeb, G. B. & Ahmad, K. Genome-wide profiling of salt fractions maps physical properties of chromatin. *Genome Res* **19**, 460-469, doi:10.1101/gr.087619.108 (2009).

310 Prigge, J. R. & Schmidt, E. E. Interaction of protein inhibitor of activated STAT (PIAS) proteins with the TATA-binding protein, TBP. *J Biol Chem* **281**, 12260-12269, doi:10.1074/jbc.M510835200 (2006).
311 Kessler, R. *et al.* dDsk2 regulates H2Bub1 and RNA polymerase II pausing at dHP1c complex target genes. *Nat Commun* **6**, 7049, doi:10.1038/ncomms8049 (2015).
312 Jiang, N., Emberly, E., Cuvier, O. & Hart, C. M. Genome-wide mapping of boundary element-associated factor (BEAF) binding sites in Drosophila melanogaster links BEAF to transcription. *Mol Cell Biol* **29**, 3556-3568, doi:10.1128/MCB.01748-08 (2009).
313 Ramírez, F. *et al.* High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun* **9**, 189, doi:10.1038/s41467-017-02525-w (2018).
314 Maudlin, I. E. & Beggs, J. D. Spt5 modulates co-transcriptional spliceosome assembly in. *RNA*, doi:10.1261/rna.070425.119 (2019).
315 Blackledge , N. P. *et al.* (bioRxiv, https://doi.org/10.1101/667667, 2019).
316 Ninova, M. *et al.* (pre-print https://doi.org/10.1101/533091, 2019).
317 Wang, M. *et al.* ME31B globally represses maternal mRNAs by two distinct mechanisms during the. *Elife* **6**, doi:10.7554/eLife.27891 (2017).
318 Igreja, C. & Izaurralde, E. CUP promotes deadenylation and inhibits decapping of mRNA targets. *Genes Dev* **25**, 1955-1967, doi:10.1101/gad.17136311 (2011).
319 Meyer, W. J. *et al.* Overlapping functions of argonaute proteins in patterning and morphogenesis of Drosophila embryos. *PLoS Genet* **2**, e134, doi:10.1371/journal.pgen.0020134 (2006).
320 Bainbridge, S. P. & Bownes, M. Staging the metamorphosis of Drosophila melanogaster. *J Embryol Exp Morphol* **66**, 57-80 (1981).
321 Ni, J. Q., Liu, L. P., Hess, D., Rietdorf, J. & Sun, F. L. Drosophila ribosomal proteins are associated with linker histone H1 and suppress gene transcription. *Genes Dev* **20**, 1959-1973, doi:10.1101/gad.390106 (2006).
322 Antão, J. M., Mason, J. M., Déjardin, J. & Kingston, R. E. Protein landscape at Drosophila melanogaster telomere-associated sequence repeats. *Mol Cell Biol* **32**, 2170-2182, doi:10.1128/MCB.00010-12 (2012).
323 Almedawar, S., Colomina, N., Bermúdez-López, M., Pociño-Merino, I. & Torres-Rosell, J. A SUMO-dependent step during establishment of sister chromatid cohesion. *Curr Biol* **22**, 1576-1581, doi:10.1016/j.cub.2012.06.046 (2012).
324 Wan, J., Subramonian, D. & Zhang, X. D. SUMOylation in control of accurate chromosome segregation during mitosis. *Curr Protein Pept Sci* **13**, 467-481 (2012).
325 Mukhopadhyay, D. & Dasso, M. The fate of metaphase kinetochores is weighed in the balance of SUMOylation during S phase. *Cell Cycle* **9**, 3194-3201, doi:10.4161/cc.9.16.12619 (2010).
326 Mukhopadhyay, D., Arnaoutov, A. & Dasso, M. The SUMO protease SENP6 is essential for inner kinetochore assembly. *J Cell Biol* **188**, 681-692, doi:10.1083/jcb.200909008 (2010).
327 Fu, H. *et al.* SENP6-mediated M18BP1 deSUMOylation regulates CENP-A centromeric localization. *Cell Res* **29**, 254-257, doi:10.1038/s41422-018-0139-y (2019).

328 Lehembre, F. *et al.* Covalent modification of the transcriptional repressor tramtrack by the ubiquitin-related protein Smt3 in Drosophila flies. *Mol Cell Biol* **20**, 1072-1082, doi:10.1128/mcb.20.3.1072-1082.2000 (2000).
329 Wakiyama, M., Imataka, H. & Sonenberg, N. Interaction of eIF4G with poly(A)-binding protein stimulates translation and is critical for Xenopus oocyte maturation. *Curr Biol* **10**, 1147-1150, doi:10.1016/s0960-9822(00)00701-6 (2000).
330 Khayachi, A. *et al.* Sumoylation regulates FMRP-mediated dendritic spine elimination and maturation. *Nat Commun* **9**, 757, doi:10.1038/s41467-018-03222-y (2018).
331 Nakamura, A., Sato, K. & Hanyu-Nakamura, K. Drosophila cup is an eIF4E binding protein that associates with Bruno and regulates oskar mRNA translation in oogenesis. *Dev Cell* **6**, 69-78 (2004).
332 Nelson, M. R., Leidal, A. M. & Smibert, C. A. Drosophila Cup is an eIF4E-binding protein that functions in Smaug-mediated translational repression. *EMBO J* **23**, 150-159, doi:10.1038/sj.emboj.7600026 (2004).
333 Safaee, N. *et al.* Interdomain allostery promotes assembly of the poly(A) mRNA complex with PABP and eIF4G. *Mol Cell* **48**, 375-386, doi:10.1016/j.molcel.2012.09.001 (2012).
334 Zhang, Y., Ling, J., Yuan, C., Dubruille, R. & Emery, P. A role for Drosophila ATX2 in activation of PER translation and circadian behavior. *Science* **340**, 879-882, doi:10.1126/science.1234746 (2013).
335 Lee, J. *et al.* LSM12 and ME31B/DDX6 Define Distinct Modes of Posttranscriptional Regulation by ATAXIN-2 Protein Complex in Drosophila Circadian Pacemaker Neurons. *Mol Cell* **66**, 129-140.e127, doi:10.1016/j.molcel.2017.03.004 (2017).
336 Jongjitwimol, J. *et al.* The S. pombe translation initiation factor eIF4G is Sumoylated and associates with the SUMO protease Ulp2. *PLoS One* **9**, e94182, doi:10.1371/journal.pone.0094182 (2014).
337 Park, E. H., Zhang, F., Warringer, J., Sunnerhagen, P. & Hinnebusch, A. G. Depletion of eIF4G from yeast cells narrows the range of translational efficiencies genome-wide. *BMC Genomics* **12**, 68, doi:10.1186/1471-2164-12-68 (2011).
338 Formstecher, E. *et al.* Protein interaction mapping: a Drosophila case study. *Genome Res* **15**, 376-384, doi:10.1101/gr.2659105 (2005).
339 Valadez-Graham, V. *et al.* XNP/dATRX interacts with DREF in the chromatin to regulate gene expression. *Nucleic Acids Res* **40**, 1460-1474, doi:10.1093/nar/gkr865 (2012).
340 Leung, J. W. *et al.* Alpha thalassemia/mental retardation syndrome X-linked gene product ATRX is required for proper replication restart and cellular resistance to replication stress. *J Biol Chem* **288**, 6342-6350, doi:10.1074/jbc.M112.411603 (2013).
341 Belotserkovskaya, R. & Reinberg, D. Facts about FACT and transcript elongation through chromatin. *Curr Opin Genet Dev* **14**, 139-146, doi:10.1016/j.gde.2004.02.004 (2004).
342 Kolybaba-Stewart, A. *The Drosophila Homologues of the Human ATRX Gene Underlying the ATRX Mental Retardation Syndrome Have Independent Functions At Chromatin* PhD thesis, Ludwig-Maximilians-Universität München, (2016).

343 Rees, J. S. *et al.* In vivo analysis of proteomes and interactomes using Parallel Affinity Capture (iPAC) coupled to mass spectrometry. *Mol Cell Proteomics* **10**, M110.002386, doi:10.1074/mcp.M110.002386 (2011).
344 Raffa, G. D., Cenci, G., Siriaco, G., Goldberg, M. L. & Gatti, M. The putative Drosophila transcription factor woc is required to prevent telomeric fusions. *Mol Cell* **20**, 821-831, doi:10.1016/j.molcel.2005.12.003 (2005).
345 Font-Burgada, J., Rossell, D., Auer, H. & Azorín, F. Drosophila HP1c isoform interacts with the zinc-finger proteins WOC and Relative-of-WOC to regulate gene expression. *Genes Dev* **22**, 3007-3023, doi:10.1101/gad.481408 (2008).
346 Sharrocks, A. D. PIAS proteins and transcriptional regulation--more than just SUMO E3 ligases? *Genes Dev* **20**, 754-758, doi:10.1101/gad.1421006 (2006).
347 Cook, P. R. & Marenduzzo, D. Transcription-driven genome organization: a model for chromosome structure and the regulation of gene expression tested through simulations. *Nucleic Acids Res* **46**, 9895-9906, doi:10.1093/nar/gky763 (2018).
348 Cook, P. R. The organization of replication and transcription. *Science* **284**, 1790-1795, doi:10.1126/science.284.5421.1790 (1999).
349 Cook, P. R. Nongenic transcription, gene regulation and action at a distance. *J Cell Sci* **116**, 4483-4491, doi:10.1242/jcs.00819 (2003).
350 Matunis, M. J., Zhang, X. D. & Ellis, N. A. SUMO: the glue that binds. *Dev Cell* **11**, 596-597, doi:10.1016/j.devcel.2006.10.011 (2006).
351 Kato, H. *et al.* RNA polymerase II is required for RNAi-dependent heterochromatin assembly. *Science* **309**, 467-469, doi:10.1126/science.1114955 (2005).
352 Bérubé, N. G., Smeenk, C. A. & Picketts, D. J. Cell cycle-dependent phosphorylation of the ATRX protein correlates with changes in nuclear matrix and chromatin association. *Hum Mol Genet* **9**, 539-547, doi:10.1093/hmg/9.4.539 (2000).
353 Clynes, D. *et al.* ATRX dysfunction induces replication defects in primary mouse cells. *PLoS One* **9**, e92915, doi:10.1371/journal.pone.0092915 (2014).
354 Koch, C. M., Honemann-Capito, M., Egger-Adam, D. & Wodarz, A. Windei, the Drosophila homolog of mAM/MCAF1, is an essential cofactor of the H3K9 methyl transferase dSETDB1/Eggless in germ line development. *PLoS Genet* **5**, e1000644, doi:10.1371/journal.pgen.1000644 (2009).
355 Riddle, N. C. *et al.* Enrichment of HP1a on Drosophila chromosome 4 genes creates an alternate chromatin structure critical for regulation in this heterochromatic domain. *PLoS Genet* **8**, e1002954, doi:10.1371/journal.pgen.1002954 (2012).
356 Li, H. *et al.* The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. *J Biol Chem* **281**, 19489-19500, doi:10.1074/jbc.M513249200 (2006).
357 Karimi, M. M. *et al.* DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell* **8**, 676-687, doi:10.1016/j.stem.2011.04.004 (2011).

358 Meehan, R. R., Kao, C. F. & Pennings, S. HP1 binding to native chromatin in vitro is determined by the hinge region and not by the chromodomain. *EMBO J* **22**, 3164-3174, doi:10.1093/emboj/cdg306 (2003).

359 Zhao, T., Heyduk, T. & Eissenberg, J. C. Phosphorylation site mutations in heterochromatin protein 1 (HP1) reduce or eliminate silencing activity. *J Biol Chem* **276**, 9512-9518, doi:10.1074/jbc.M010098200 (2001).

360 Hsieh, T.-H. S. *et al.* (https://www.biorxiv.org/content/10.1101/638775v1, 2019).

361 Liu, H. W., Banerjee, T., Guan, X., Freitas, M. A. & Parvin, J. D. The chromatin scaffold protein SAFB1 localizes SUMO-1 to the promoters of ribosomal protein genes to facilitate transcription initiation and splicing. *Nucleic Acids Res* **43**, 3605-3613, doi:10.1093/nar/gkv246 (2015).

362 Ashburner, M. Drosophila: A Laboratory Handbook and Manual. Two volumes. (1989).